Biochemical and computational studies towards selective inhibition of the immunoproteasome

A thesis submitted to Middlesex University in partial fulfilment of the requirements for the degree of:

Master of Science (by Research)

Duncan Jamie Allardyce

Student number: M00551417

Supervisor - Dr. Erika Loizidou
Director of studies - Dr. Celia Bell

Department of Natural Sciences
Middlesex University

September 2017

Abstract

Biochemical and computational studies towards selective inhibition of the immunoproteasome.

Allardyce, D.J., Bell, C. and Loizidou, E.

Department of Natural Sciences, Middlesex University, London, NW4 4BT.

The proteasome pathway degrades >90% of cytosolic proteins deemed redundant, misfolded or toxic, thereby influencing key regulatory pathways including: cell cycle control, DNA repair and apoptosis. As such, proteasome inhibitors (PI) have exhibited broad therapeutic applications, particularly for multiple myeloma and mantle cell lymphoma with 3 inhibitors gaining FDA approval. However, covalent binding and lack of targeted action cause severe toxicity. Upon stimulation by inflammatory cytokines, constitutive proteasome (CP) active sites β1c, β2c and β5c are replaced with corresponding β1i, β2i and β5i subunits; forming the immunoproteasome (IP). The abundant CP is required for regular cell function, however due to upregulation in diseased states selective IP inhibition is associated with an increased therapeutic index. Recent identification of structural differences between CP and IP specificity pockets (S1-4) allows structure-based drug design.

The cyclic peptides argyrin A and F exhibit potent, reversible CP inhibition with mechanisms distinct to existing therapeutics. In this project, argyrin B inhibition and binding interactions between the CP and IP are investigated, using purified enzyme assays alongside computational molecular modelling. Kinetic assays revealed argyrin B IC_{50} values of 146.5 µM and 8.76 µM at β1c and β1i, respectively; a 16-fold difference with statistical significance. Whilst argyrin B also showed slight preference towards β5i over β5c, with low micromolar IC_{50} values. The same trends were supported by K_i values and molecular docking estimated binding energies. AutoDock and FRED simulations suggest increased β1i S1 pocket hydrophobicity, T21S and G97H substitutions from β1c to β1i as key towards favourable β1i binding. At β5c, small, hydrophobic characteristics of S2 become polar in β5i that enhances argyrin B interactions. These findings facilitate design of further IP selective inhibitors, whilst the identification of the first known β1i selective and non-covalent PI shows great therapeutic potential with reduced toxicity proposed in comparison to existing therapeutics.

Acknowledgments

I would like to thank Middlesex University and particularly Dr. Celia Bell for providing me the opportunity to undertake this Masters by Research degree. I appreciate the time and support granted to make this possible. Thank you to Dr. Erika Loizidou who created this brilliant project and provided the expertise to progress the research. I am grateful for all of the training, advice and guidance throughout. I have always been able to count on immediate answers to questions and provision of quality feedback. Further thanks to my wife, family, friends and colleagues for continued support. Finally, thanks go to Novartis for the donation of argyrin B compound.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

ADME	Absorption distribution metabolism and excretion
AICc	Akaike's information criterion
Ala	Alanine
AMBER	Assisted model building with energy refinement
AMC	7-amino-4-methylcoumarin
AMP	Adenosine monophosphate
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATG	Autophagy protein
ATP	Adenosine triphosphate
BAX	Bcl-2-associated X protein
Bcl	B-cell lymphoma
BLAST	Basic local alignment search tool
BrAAP	Branched amino acid preferring peptidase
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CHOP	CCAAT/Enhances-binding protein homologous protein
CP	Constitutive Proteasome
Cys	Cysteine
DMSO	Dimethyl sulphoxide
DTT	Dithiothreitol
DUB	Deubiquitinating enzyme
E	Enzyme
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme
E3	Ubiquitin-protein ligase enzyme
EDTA	Ethylenediaminetetraacetic acid
EIF2a	Eukaryotic translation initiation factor 2A
ELISA	Enzyme linked immuno sorbent assay
ER	Endoplasmic reticulum
FADD	Fas-associated death domain
FDA	Food and Drug Administration
FEB	Free energy perturbation
GFP	Green fluorescent protein

Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GROMACS	Groningen machine for chemical simulations
His	Histidine
I	Inhibitor
IκB	Inhibitory kappa B
IAP	Inhibitor of apoptosis
IFN-γ	Interferon gamma
Ig	Immunoglobulin
Ile	Isoleucine
IP	Immunoproteasome
K_d	Dissociation constant
K_i	Inhibition constant
$K_{ic/u}$	Competitive / un-competitive inhibition constant
K_m	Michaelis-Menten constant
Leu	Leucine
LigParGen	Ligand parameter generator
LMP	Large multifunctional peptidase
Lys	Lysine
MAPK	Microtubule associated protein kinase
MCL	Mantle cell lymphoma
MECL	Multicatalytic endopeptidase complex-like
Met	Methionine
MHC	Major histocompatibility complex
MM	Multiple Myeloma
NBS	Non-binding surface
NF-kB	Nuclear factor kappa B
P	Product
PI	Proteasome inhibitor
PA28	Proteasome activator 28
PAGE	Polyacrylamide gel electrophoresis
PAM	Point accepted mutation
PDB	Protein data bank
Phe	Phenylalanine

POMP	proteasome maturation protein
Pro	Proline
PSMB	Proteasome subunit beta
R	Ideal gas constant
RAGE	advanced-glycation-end-product receptor
RMSD	Root-mean-square deviation
Rpn	Regulatory particle of non-ATPase
Rpt	Regulatory particle of triple-ATPase
S	Substrate
S1-4	Specificity pocket 1-4
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEM	Standard error of mean
Ser	Serine
SMAC	Second mitochondria-derived activator of caspases
SNLR	Simultaneous non-linear regression
STRENDA	Standards for reporting enzyme data
Suc	Succinyl
Thr	Threonine
TNF-α	Tumour necrosis factor alpha
TP	Thymoproteasome
TRAPP	Transient pockets in proteins
Trp	Tryptophan
Tyr	Tyrosine
Ub	Ubiquitin
V_{app}	Apparent maximal velocity
V_{max}	Maximal velocity
Val	Valine
VEGF	Vascular endothelial growth factor

Chapter 1. Introduction

1.1 Protein degradation

1.1.1 Proteostasis

The role of protein degradation is vital for cellular growth and viability. Degradation rate controls the half-life of proteins, thereby influencing intracellular protein levels to maintain proteostasis. In eukaryotic cells, there are two major pathways that control protein degradation; lysosomal proteolysis and the ubiquitin-proteasome system (Etlinger and Goldberg, 1977).

Lysosomes contain a concoction of digestive enzymes such as proteases and cathepsins, which allow the degradation of proteins that have been engulfed through endocytosis and autophagy. This process is generally non-selective; however under nutrient deprivation, degradation of specific proteins can serve as an energy supply (Cuervo and Dice, 1996). Lysosomes play a vital role in immune response through presentation of peptide fragments on immune cells, via major histocompatibility complex (MHC) class II receptors. The pathway was originally identified as the main process of proteolysis resulting in the hydrolysis of diverse macromolecules.

However, in 1977, the nonlysosomal, ubiquitin-proteasome pathway was first introduced by Etlinger and Goldberg (1977), presenting a role in degradation of polyubiquitin tagged proteins that are deemed redundant, misfolded or potentially toxic (Goldberg, 2003). This selective degradation system is present in all eukaryotic cells and is responsible for >90% of intracellular degradation. Key regulatory proteins and molecular pathways are affected such as: cell cycle control, DNA repair, differentiation, stress response, antigen presentation and apoptosis (Pellom and Shanker, 2012). With greater understanding of the molecular mechanisms involved, the proteolytic pathway was soon linked with several disease states and has since become a long standing therapeutic target. The proteasome complex, found within the nucleus and cytoplasm and constituting approximately 0.5%-0.8% of total cellular protein itself (Parlati *et al.*, 2009), is the machinery of nonlysosomal protein degradation.

1.1.2 Polyubiquitylation

The proteasome architecture of regulatory caps determines the type of proteins recognised and mechanism of action, using polyubiquitylation as the most common pathway.

Specific conjugation of ubiquitin (Ub) with distinct length and linkage, tags the protein targets for recognition and rapid degradation by the proteasome. The ubiquitination process, summarised in figure 1, involves activation, conjugation and ligation steps, facilitated by E1, E2 and E3 enzymes respectively. Adenosine triphosphate (ATP) is required for the attachment of ubiquitin as well as 26S proteasome actions.

Firstly, ubiquitin is activated by E1, using ATP and is transferred to E2. E3 recognises and binds the specific substrate based on primary protein sequences, therefore accounting for substrate selectivity. E3 is able to transfer the ubiquitin to covalently bind the substrate, generally at a lysine residue, thereby tagging for recognition by the 19S cap (Glickman and Ciechanover, 2002). The polyubiquitination mechanism remains contested, with either elongation by sequential addition of ubiquitin, or ready-formed poly-Ub chains wholly added to the protein (Pellom and Shanker, 2012).

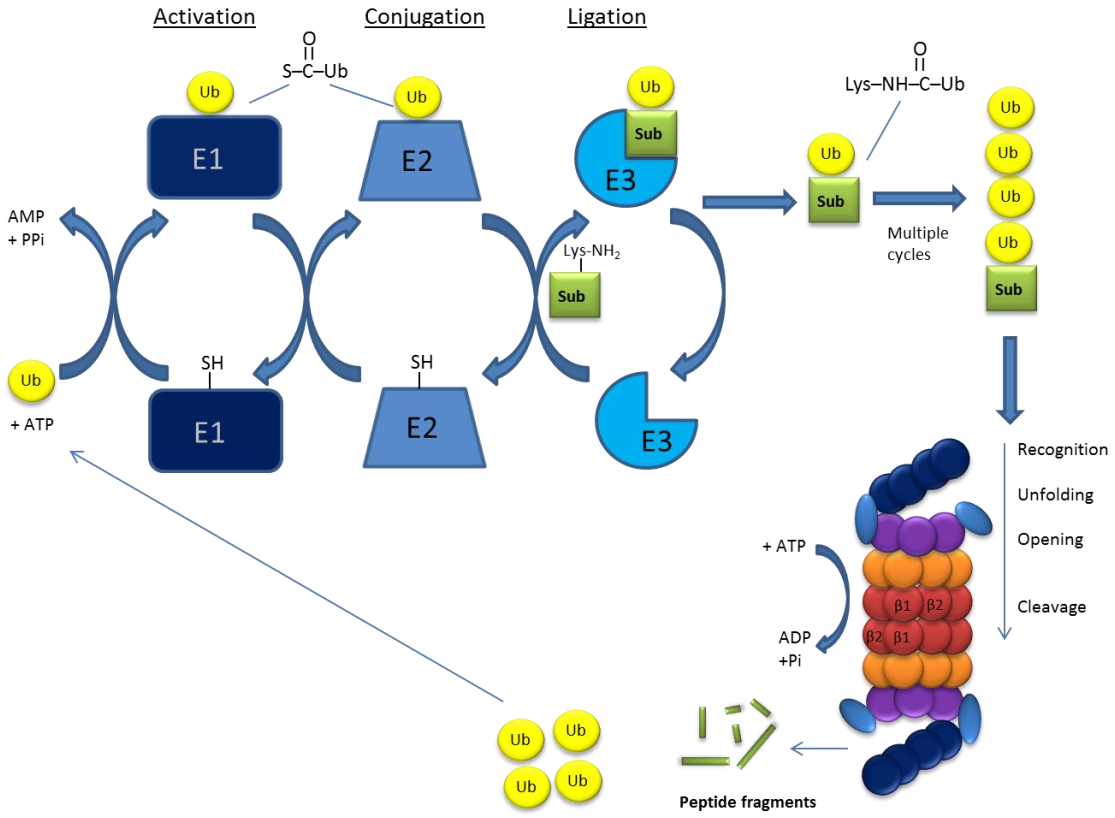


Figure 1 - Summary of the ubiquitin degradation process – Ubiquitin (Ub) undergoes 3 key steps of: activation, conjugation and ligation at E1, 2 and 3 enzymes, respectively. Upon covalent binding of substrate to a ubiquitin chain, the substrate is tagged for recognition by the 26S proteasome complex. The proteasome regulatory cap is able to unfold and open the complex allowing cleavage within the core at β catalytic sites. Peptide fragments are produced and ubiquitin is recycled.

Once the protein is signalled for degradation, the proteasome is able to recognise and process the breakdown into peptide fragments, typically of 5-25 residues in length. Before processing in the complex structures of the 26S proteasome, proteins are de-ubiquitinated by deubiquitinating enzymes (DUBs) and unfolded by the 19S cap, using ATP (Glickman and Ciechanover, 2002).

1.2 The proteasome

1.2.1 Proteasome structure

The 26S proteasome is a large (150x115 Å), cylindrical complex comprised of approximately 66 proteins arranged in a 20S catalytic core and 2 19S regulatory caps at either end, as shown in figure 2 (Pellom and Shanker, 2012). The 19S regulatory proteins are each comprised of base and lid structures, containing at least 19 subunits, with key proteins highlighted in figure 2, that facilitate the opening of the proteasome and unfolding of proteins (Huang *et al.*, 2016). The 20S core contains 2 stacked, heteroheptameric catalytic β1-7 subunit rings (figure 2, red spheres) between 2 structural α1-7 subunit rings (figure 2, orange spheres).

The constitutive proteasome (CP) is required for regular cell function with proteolytic activity occurring within β1, 2 & 5 active sites. Figure 3 shows substitution of active sites to create the immunoproteasome (IP) and thymoproteasome (TP) variants. IP formation is stimulated by inflammatory signals and bears the same function, with slightly different active sites (Huber et al., 2012). The IP is therefore upregulated in diseased states and not commonly found in healthy cells. The TP is significantly less abundant, found only in cortical thymic epithelial cells and plays a role in the adaptive immune system (Nitta *et al.*, 2010). However, it is also important to note that mixed, intermediate proteasomes containing a split of both CP and IP active sites have been reported in a variety of cell types, at widely varying levels up to an abundance of 50% (Guillaume *et al.*, 2010).

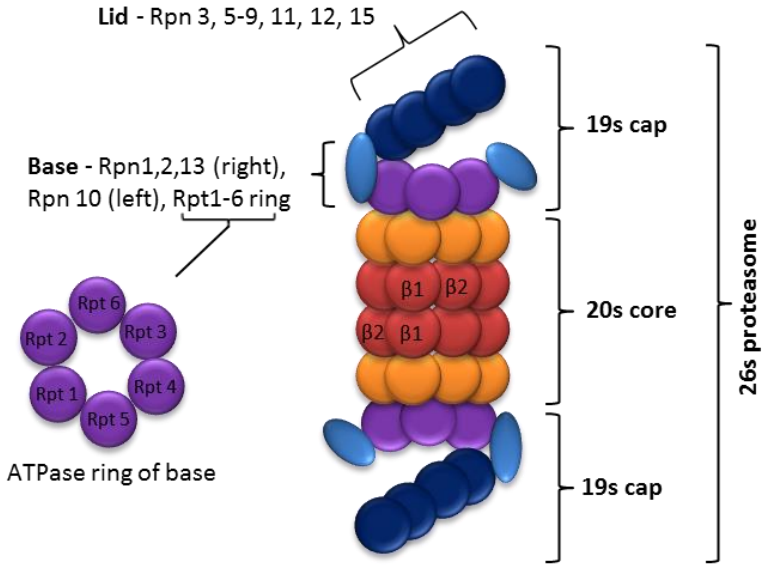


Figure 2 - Representation of mammalian 26S proteasome - Displaying 20S core comprised of heptameric α (orange) and β (red) rings, with β-subunit active sites. The 19S cap consists of lid and base structures for recognition and unfolding of substrates. The lid is comprised of regulatory particle of non-ATPase (Rpn) subunits, whilst base consists of select Rpn subunits and a regulatory particle of triple-ATPase (Rtp) subunits ring.

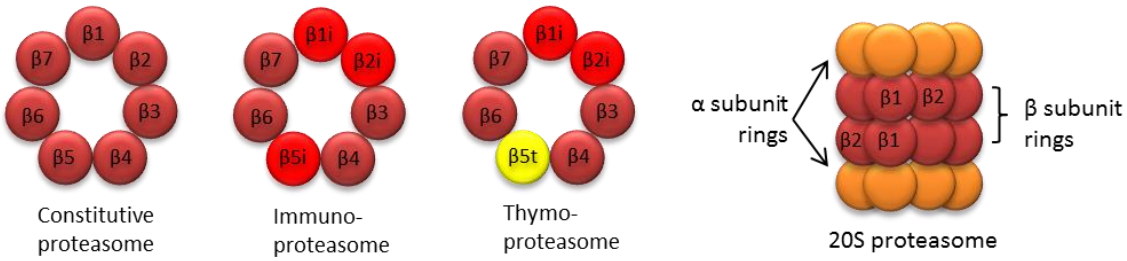


Figure 3 - Illustration of 20S proteasome variants – Constitutive proteasome, immunoproteasome and thymoproteasome structures are shown with substituted β-active site subunits in each β-ring. α-subunit rings remain unchanged.

The human 20S proteasome (PDB: 4R30) was first crystallised in 2015 by Harshbarger *et al.* (2015) using x-ray diffraction, to a resolution of 2.6 Å. The characterisation of active sites and recognition proteins allows further understanding of molecular mechanisms involved in the degradation process. During protein recognition and degradation, 3 dynamic conformations of the 26S proteasome have been proposed: accepting substrates, an intermediate positioning and translocation to the 20S particle (Huang *et al.*, 2016). These dynamic conformations facilitate

understanding of protein recognition and control of peptides to be degraded within the catalytic core.

Upon reaching the 20S subunit, there are 3 main active sites: β1, β2 and β5 subunits, with distinct mechanisms of action that allow for levels of specificity as discussed in the following sections.

1.2.2 Threonine protease activity mechanism

The active sites work synergistically to cleave proteins using the same mechanism; however each displays distinct substrate specificities on account of the interactions at specificity pockets surrounding the active site. After series of mutation studies, it was found that the N-terminal threonine 1 (Thr1) residues at the beginning of β1, β2 and β5 subunit chains provide the site for proteolytic activity. Despite β7 also displaying an N-terminal Thr residue, its' removal results in no change to activity, suggesting that this is catalytically inactive (Dick et al., 1998). The threonine acts as a nucleophile hydrolase in an endoprotease reaction to cleave the scissile bond of substrates, as summarised in figure 4.

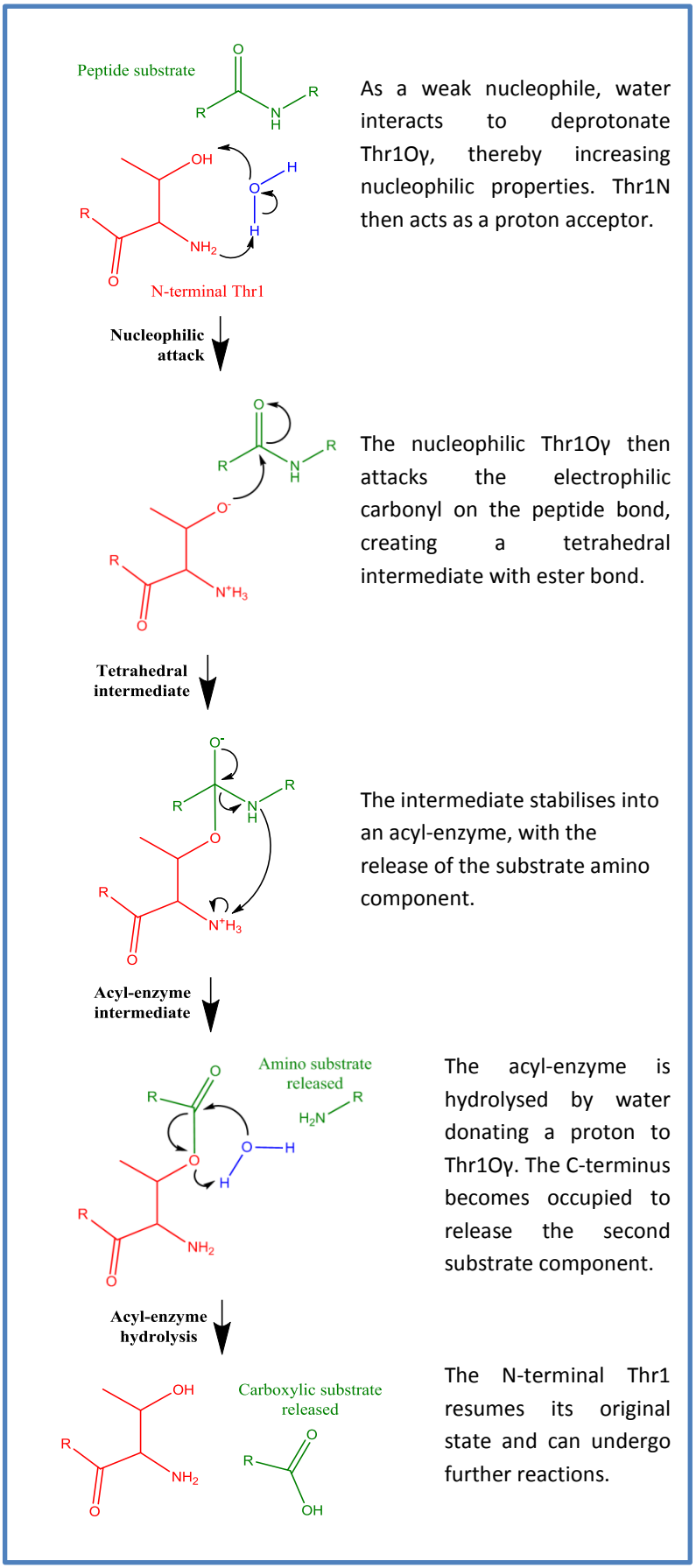


Figure 4 - Schematic representation of substrate proteolysis mechanism - The N-terminal threonine residue, shown in red, is the catalytic centre within each proteasome active site β-subunit. The substrate amide bond is shown in green and water molecules in blue. (Adapted from; Kisselev and Goldberg, 2001).

Understanding of the proteolysis mechanism can facilitate drug design of the target. Inhibitors with an electrophilic group to react with the nucleophilic N-terminal Thr1Oy residue at active sites are established as effective whilst other modes of inhibition are also possible (Kisselev et al., 2012).

1.2.3 Active sites and specificity pockets

The surrounding residues of Thr1 create specific pockets that contribute to the selectivity and stabilisation of substrates. The specificity binding pockets S1-3 accommodate the amino acid side chains, as shown in figure 5 (Huber *et al.*, 2012). Due to amino acid variation between β1, β2 and β5, the characteristic activity of each active site differs.

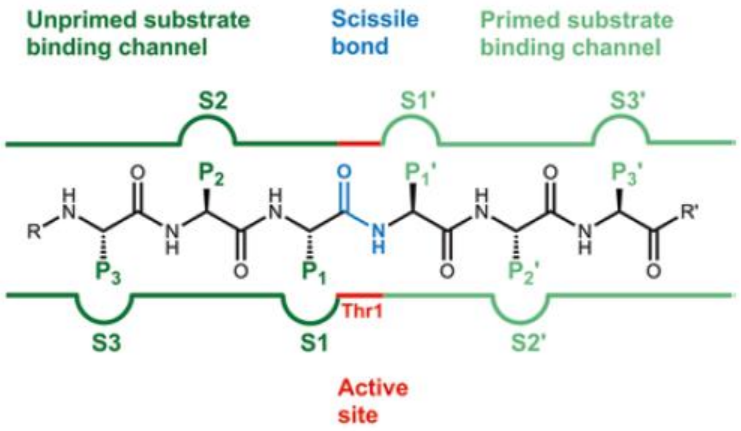


Figure 5 - Proteasome active site substrate binding channels – Within each β active site, amino acid side chains P1-3 of the substrate, bind specificity pockets S1-3. Specificity pockets each have characteristic shapes and properties that produce preference for particular substrates. Upon binding, the Thr1 residue cleaves the scissile peptide bond of the protein (Huber *et al.*, 2012).

β1, β2 and β5 are characterised with basic, acidic and nonpolar specificity pockets, respectively. As such, β1 displays caspase-like activity with cleavage after acidic amino acids, β2 cleaves after basic amino acids with trypsin-like activity and β5 cleaves after hydrophobic residues in a chymotrypsin-like activity (Dick *et al.*, 1998). These properties are summarised in table 1 below and visualised in figure 6.

Table 1 - CP active site characteristics – Each active site is termed according to its general proteolysis activity, it does not perfectly represent the sole mechanism. $^{\bullet}$ indicates quicker than glutamate, * is not applicable if followed by proline, $^{\phi}$ denotes decreased cleavage rates. (Adapted from; Dick *et al.*, 1998).

Active Site	Activity	Cleaves after	
β1	Caspase-like	Acidic amino acids	Asp$^{\bullet}$, Glu
β2	Trypsin-like	Basic amino acids	Arg, Lys
β5	Chymotrypsin-like	Hydrophobic residues	Phe, Trp, Tyr*, His$^{\phi}$, Met$^{\phi}$, Leu$^{\phi}$

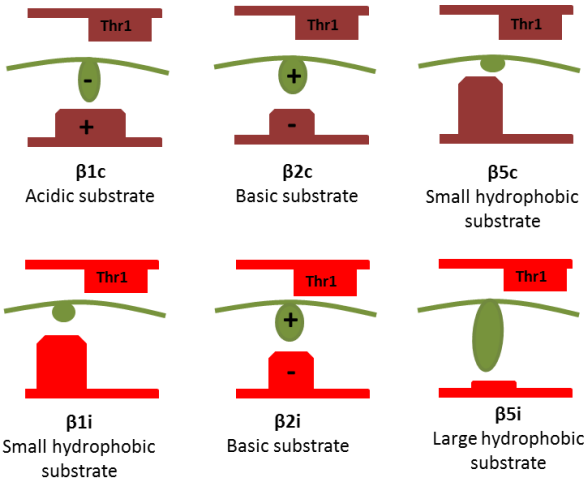


Figure 6 – Representation of characteristic size and properties of CP and IP active sites – Substrate peptide chains are displayed in green with protruding R-groups. Red represents active site surrounding catalytic Thr1, revealing differences in size and charges. Preference of substrate characteristics are highlighted within text labels.

Based upon these activities, inhibitor compounds can exhibit greater selectivity towards particular active sites. β5 was originally considered a significantly more influential active site than β1 and β2, therefore becoming a focus in early proteasome inhibitor development. However, Kisselev (2006) showed the inhibition of β5c to reduce overall protein breakdown by 11-50%, whilst β1 and β2 inhibition reduce overall activity by 12-22% and 3-35%, respectively. This shows a hugely varying impact on overall degradation, as well as the importance of each active site. In addition, there appears to be some compensatory actions since the inhibition of 2 active sites displayed a greater reduction of degradation than the sum of each individually.

As well as key differences between β1, 2 & 5, there are variations in the active subunit sequences between the CP and IP corresponding sites. This leads to different binding preferences of substrates and inhibitors. As shown in figure 6, β1i generally accepts smaller, hydrophobic

substrates as opposed to acidic preference at β1c. β2c and β2i sites are similar and β5i are able to accommodate larger hydrophobic residues than β5c.

1.3 The immunoproteasome

1.3.1 Formation and structure of the immunoproteasome

Upon stimulation by inflammatory cytokines such as interferon gamma (IFN-γ) and to a lesser extent, tumour necrosis factor alpha (TNF-α), the proteasome is phosphorylated and transcription of IP specific subunits begins. Large multifunctional peptidase 2 (*LMP2)*, multicatalytic endopeptidase complex-like-1 (*MECL-1)* and *LMP7* genes encode for β1i (proteasome subunit beta-9 - PSMB9, LMP2), β2i (PSMB10, LMP10, MECL-1) and β5i (PSMB8, LMP7), respectively which replace corresponding 20S CP subunits; reviewed by Johnston-Carey *et al.* (2015). In addition, combinations of 19S and 11S regulatory caps are formed as displayed in figure 7. This 11S structure performs the same function as 19S although it allows for non-ubiquitinated peptide degradation, improves efficiency and plays a greater role in antigen presentation (Engelhard, 1994).

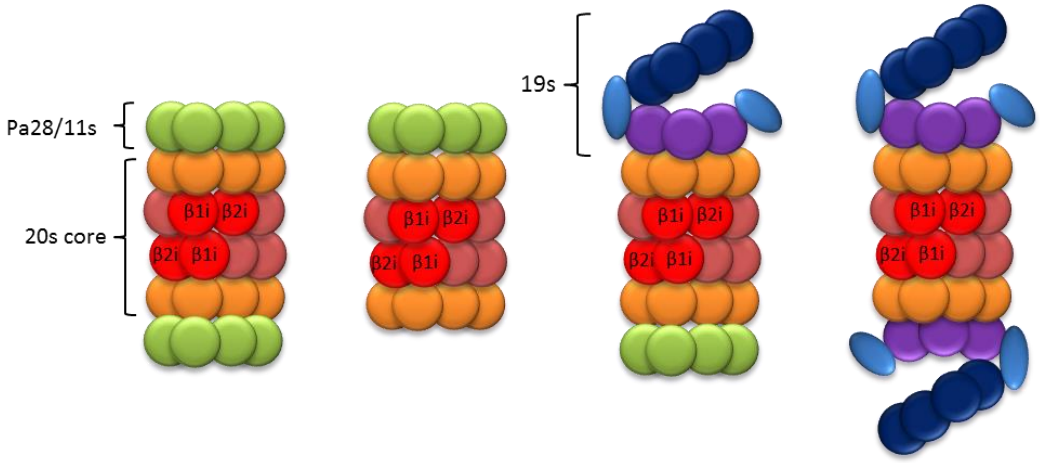


Figure 7 - Immunoproteasome structure variation – Representation of different 11S and/or 19S regulatory caps combinations on 20S core particles. The 19S regulatory cap (blue and purple) performs the same function as Pa28/11S cap (green), although the presence of Pa28/11S also allows for non-ubiquitinated peptide degradation and increased rates.

Under standard conditions, the levels of IP are low in comparison to CP, however, during inflammatory and diseased states, levels of IFN-γ and oxidative stress are increased, resulting in elevated proportions of IP. Interestingly the IP is proposed to exhibit approximately double the activity of the CP and a half-life of 27 hours, compared to 133 hours of the CP (Heink *et al.*, 2005). IFN-γ is able to upregulate proteasome maturation protein (POMP) expression that increases the speed of formation as well as preference to IP subunits over CP (Gregerson and Ferrington, 2012). This suggests that the IP is upregulated upon demand, for rapid degradation in select tissues before returning to normal states. Indeed, there are high levels of IP reported in inflammatory and autoimmune disorders, diseases of the central nervous system and some cancers (Miller *et al.*, 2013).

Although largely conserved, there are slight differences between the amino acid chain and therefore active site structure of the CP and IP corresponding active sites. Figure 8 compares a surface representation view based around the catalytic Thr1 (red), of human β1c (blue) compared to humanised β1i (green), with visible changes circled. These may lead to different substrate or compound binding preferences.

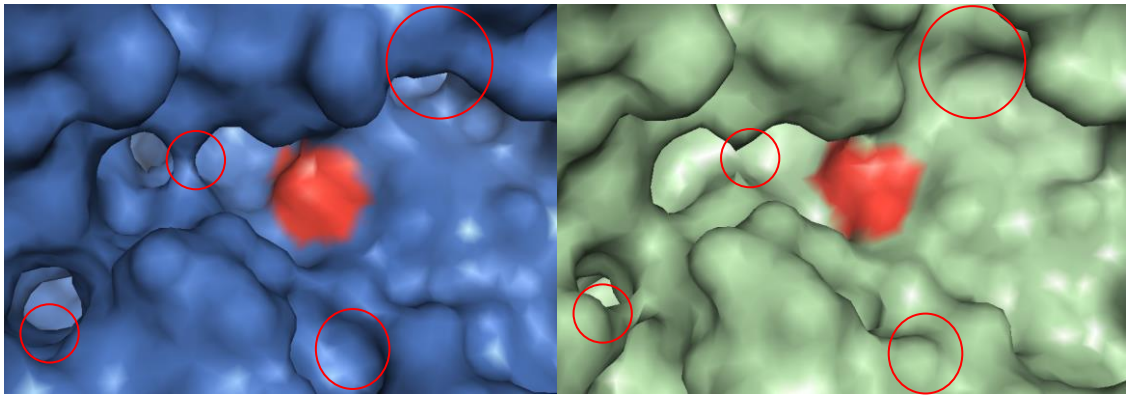


Figure 8 - Structural differences between CP and IP β1 active site - Surface representation of human β1c (blue) and humanised β1i (green), with threonine-1 catalytic centre in red. Significant structural differences are circled.

1.3.2 Function of the immunoproteasome

The IP is shown to play a vital role in immunity, providing peptides for MHC antigen presentation that differ from those the CP produces. Whilst β2i and β5i share the characteristic trypsin-like and chymotrypsin-like activities of their CP counterparts, β1i exhibits trypsin-like activity as opposed to caspase-like of β1c as figure 6 displays. This contributes towards the IP's overall increased

chymotrypsin-like activity compared to the CP (Pellom and Shanker, 2012; Johnston-Carey et al., 2015). The absence of β5 was shown to lower MHC class I expression by up to 50%, whilst β1 and β2 showed no impact. This led to the discovery that the hydrophobic residues produced by chymotrypsin-like activity, bind with greater affinity to MHC class I molecules (Engelhard, 1994; Nitta *et al.*, 2010).

In addition, the TP composed of β1i, β2i and β5t subunits as shown in figure 3 plays a vital role in immunity. Interestingly, the β5t shows lower chymotrypsin-like activity compared to constitutive and immune counterparts. Despite this, the TP is required for production of hydrophobic residues to facilitate development of cytotoxic CD8+ T cells (Nitta *et al.*, 2010).

As well as established adaptive immune function, the IP has been shown to regulate activation of alveolar macrophages, revealing a novel role in the innate immune response. β5i specific inhibition is proposed as a suitable approach for pneumonia and sepsis related acute respiratory distress syndrome (Chen *et al.*, 2016).

Beyond immune functions, the IP is linked with an increasing variety of disease states, later discussed in section 1.4.3 and provides a promising avenue for more targeted treatment on account of their distribution in types of cells.

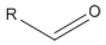
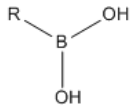
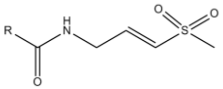
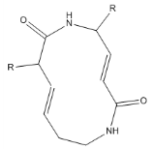
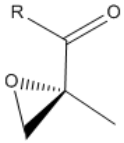
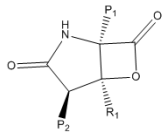
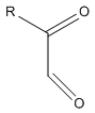
1.4 The proteasome as a drug target

1.4.1 Classes of proteasome inhibitors

It has been reported that approximately half of existing drugs target enzymes (Copeland, 2005). The CP and IP are highly influential enzymes in the progression of many diseases, making these major drug targets. Research utilising proteasome inhibitors has significant clinical benefit, as well as use in advances of understanding disease mechanisms, cell regulation, quality control systems and the immune response (Goldberg, 2012).

The proteasome, like most enzymes, is amenable to small molecule drugs that may use a range of mechanistic actions. Most inhibitors contain a peptide backbone that mimics the substrate with a pharmacophore that defines the class and mechanism of inhibitor. Increasingly, CP and IP inhibitors in development are also providing co-crystallised structures with the proteasome to help further elucidate mechanism of action. The main classes are summarised in table 2.

Table 2 – Major classes of proteasome inhibitors – General inhibitor classification classes alongside chemical bonds formed and associated examples of existing inhibitors.

Class	Structure	Bonding	Compounds
Aldehydes		Covalent, reversible hemiacetal adduct with N-terminal Thr1	MG132 (Rock *et al.*, 1994)
Boronates		Boron accepts oxygen lone pair of N-terminal Thr1OH, forming a stable, tetrahedral complex. Reversible, slow dissociation rates. Pro-drug with citrate ester - hydrolysis releases boronic acid	Bortezomib (PS-341, Velcade) (Groll *et al.*, 2006a) Ixazomib (MLN9708, Ninlaro) (Moreau *et al.*, 2016)
Vinyl sulphones		Covalent ether bond formed with Thr1O$^{\gamma}$	Vinyl sulfone (Kisselev *et al.*, 2012)
Vinyl ketones (syrbactins)		Covalent, irreversible ether bond Michael type 1-4 addition to Thr1OH	Syringolin A (Clerc *et al.*, 2009)
Epoxyketones		Irreversible covalent bond forms 7-membered ring adduct with Thr1N Irreversible bivalent reaction with Thr1O$^{\gamma}$ and Thr1N	Epoxomycin (Groll *et al.*, 2000) Carfilzomib (PR-171, Kyprolis) (Harshbarger *et al.*, 2015) ONX 0914 (PR-957) (Muchamuel *et al.*, 2009) Oprozomib (ONX-0912) (Parlati *et al.*, 2009) PR-924 (Parlati *et al.*, 2009)
β-lactones		Reversible or irreversible. Ester bond to Thr1 and tetrahydrofuran ring formation.	Marizomib (Salinosporamide A) (Feling *et al.*, 2003; Groll *et al.*, 2006b)
Ketoaldehydes		Reversible cyclic Schiffbase formed with Thr1O$^{\gamma}$ and Thr1N	Peptidyl glyoxal (Gräwert *et al.*, 2011)

Each class of inhibitor has benefits and limitations. From a therapeutic point of view, initially potency is of the utmost importance. However, many drugs have subsequently suffered with poor selectivity and unwanted interactions *in vivo*. For example, peptide aldehydes are broken down into inactive forms within the cell, rendering these unsuitable therapeutics, despite their potency (Pellom and Shanker, 2012). Acting upon other targets such as serine and cysteine proteases create unwanted side effects and provide an additional obstacle to overcome. Similarly, the formulation and delivery of drugs are often additional complications to overcome later in the development process (Strelow *et al.*, 2012).

As displayed in table 2, many existing proteasome inhibitors contain variations of an electrophilic warhead group, forming covalent bonds with the catalytic Thr1 of each active site. A strong chemical bond at this site can cause irreversible inhibition, associated with more severe toxicity since proteasomes are vital for regular functioning within cells (Kisselev *et al.*, 2012). Non-covalent inhibitors without a reactive group may have lower potency, although generally have fewer side effects with time limited inhibition. Rapid binding and dissociation, possible with reversible inhibitors, can also allow greater tissue distribution (Gräwert *et al.*, 2012). With synthetic motifs able to enhance interactions and affinity, non-covalent inhibitors are an emerging class of proteasome inhibitors. Upon understanding how inhibitors bind to the proteasome and active sites, it is important to elucidate the molecular pathways triggered by this event.

1.4.2 Proteasome inhibition molecular pathways

Beyond the binding mechanism, the downstream effects of proteasome inhibitors are diverse, with some examples of simplified pathways summarised in figure 9. Ultimately, many of these lead to apoptosis and cell cycle regulation.

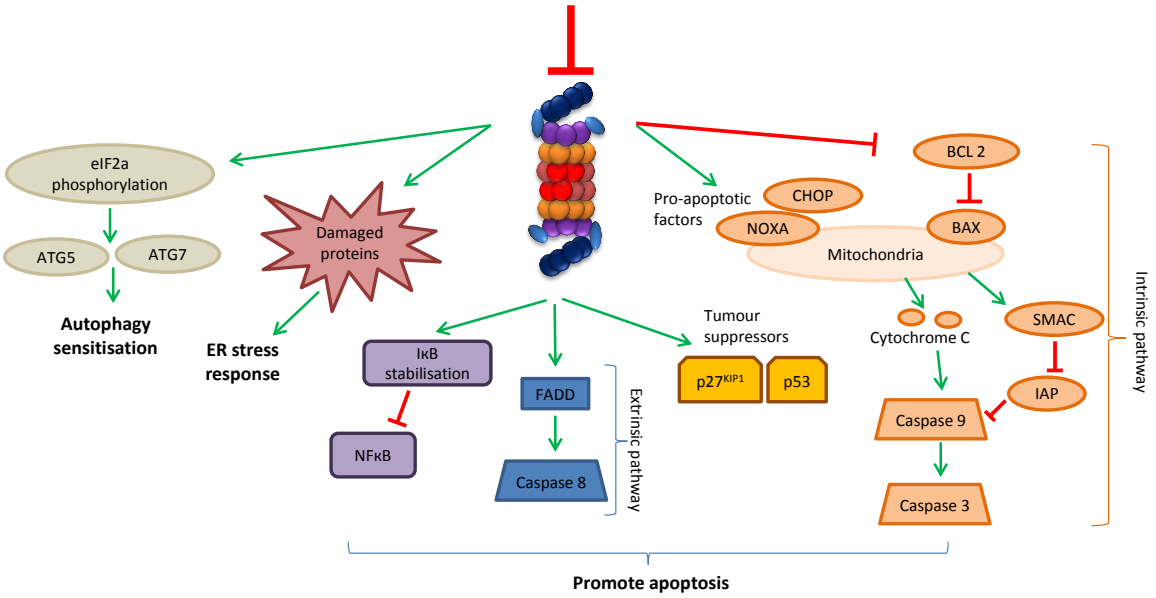


Figure 9 – Molecular pathways triggered by proteasome inhibition – Examples of key pathways and molecular changes as a result of proteasome inhibtion.

Proteasome inhibition activates autophagy in order to remove protein aggregates (Zhu *et al.*, 2010). An accumulation of damaged proteins triggers endoplasmic reticulum (ER) stress and an unfolded protein response (Pandey *et al.*, 2007). Due to the roles of the proteasome it is suggested that inhibitors reduce the activity of nuclear factor kappa B (NFkB) (Palombella *et al.*, 1994), a key transcription regulator, however this has not been consistently observed (Parlati *et al.*, 2009). Furthermore, the extrinsic apoptosis pathway is promoted through fas-associated death domain (FADD) induced caspase-8 activation (Brooks *et al.*, 2010). In addition, several pathways of intrinsic apoptosis are stimulated through a number of pro-apoptotic factors (Nickeleit *et al.*, 2008; Parlati *et al.*, 2009; Singh *et al.*, 2009). However, the complexity of the molecular mechanisms is evident by contradictory effects, such as the activation of p38 microtubule associated protein kinase (MAPK) anti-apoptotic pathways (Shi *et al.*, 2006).

It is important to note that the effector pathway varies with different cells and inhibitors. However, the broad impact provides promising avenues for drug development since differing mechanisms may allow for more targeted actions or help overcome resistance.

1.4.3 Therapeutic potential

With a greater understanding of the molecular mechanisms involved, the proteolytic pathway was soon linked with several diseases states including those shown in figure 10 and has been a long

standing therapeutic target. In particular, the proteasome's proposed role in activation of the transcription factor NF-κB (Palombella *et al.*, 1994), a key regulator in the cell cycle, kick-started research for the potential of proteasome inhibitors in cancer treatment.

Since malignant cells display characteristically rapid proliferation and genetic instability, the proteasome is a vital system required to remove any damaged proteins. Therefore inhibition of the proteasome will exhibit a more profound effect on cancer cells, compared to healthy, normal cells (Singh *et al.*, 2010). Irregular elevated levels of immunoglobulins and lymphocytes in multiple myeloma (MM) and mantle cell lymphoma (MCL) respectively make these diseases amenable to proteasome inhibitors, with 3 currently food and drug administration (FDA) approved for treatment by selectively destroying immunoglobulin-producing plasma cells (Groll *et al.*, 2006a). However, proteasome inhibitors that either reduce or stimulate cancer progression is often specific to the CP or IP as well as particular subunit upregulation or downregulation (Johnston-Carey *et al.*, 2015). IP levels are varied between different types of cancers, with little current understanding to the significance of downregulated levels. MM, lung and prostate cancer have all displayed high levels of the IP (Miller *et al.*, 2013), therefore providing opportunities for a more targeted treatment if selectivity can be achieved.

Beyond cancers, the proteasome also impacts the immune system partly because degraded peptides can act as antigens on the surface of immune cells through binding MHC class I receptors. Proteasome inhibitors therefore show potential in the treatment of autoimmune disorders such as arthritis (Muchamuel *et al.*, 2009), colitis (Basler *et al.*, 2010) and encephalomyelitis (Basler *et al.*, 2014). Furthermore, other peptide fragments generated by the proteasome would be expected to retain some biological function that is currently unknown.

More diverse applications are also under investigation such as anti-parasitics in malaria (Stadelmann *et al.*, 2014; Li *et al.*, 2016), anti-bacterials in tuberculosis (Hsu *et al.*, 2017), crop protection agents (Citovsky *et al.*, 2009), promotion of bone growth and treatment of male pattern baldness (Mundy *et al.*, 2007); highlighting the diverse and complex pathways effected. As well as clinical benefits, the development of proteasome inhibitors continues to progress the understanding of the proteasome and the cells' intracellular functions (Kisselev et al., 2006).

In disease states with increased demand for protein degradation, IP levels are raised as a coping mechanism. This phenomenon is observed in neurodegenerative diseases including Alzheimer's (Mishto *et al.*, 2006) and Huntington's (Díaz-Hernández *et al.*, 2003). IP levels can also be directly increased by the pro-inflammatory cytokine immune response to the disorder, as observed in Duchenne muscular dystrophy (Azakir *et al.*, 2014; Chen *et al.*, 2014; Farini *et al.*, 2016) as well as

cancers. Interestingly, the activation of advanced-glycation-end-product receptors (RAGE), an Alzheimer's hallmark, produces an alternative mechanism pathway for IP activation, distinct from IFN-γ (Grimm *et al.*, 2012). Due to the upregulation observed in disease states, the IP has arguably emerged as a more attractive drug target over the CP that is abundant in all cells (Miller *et al.*, 2013).

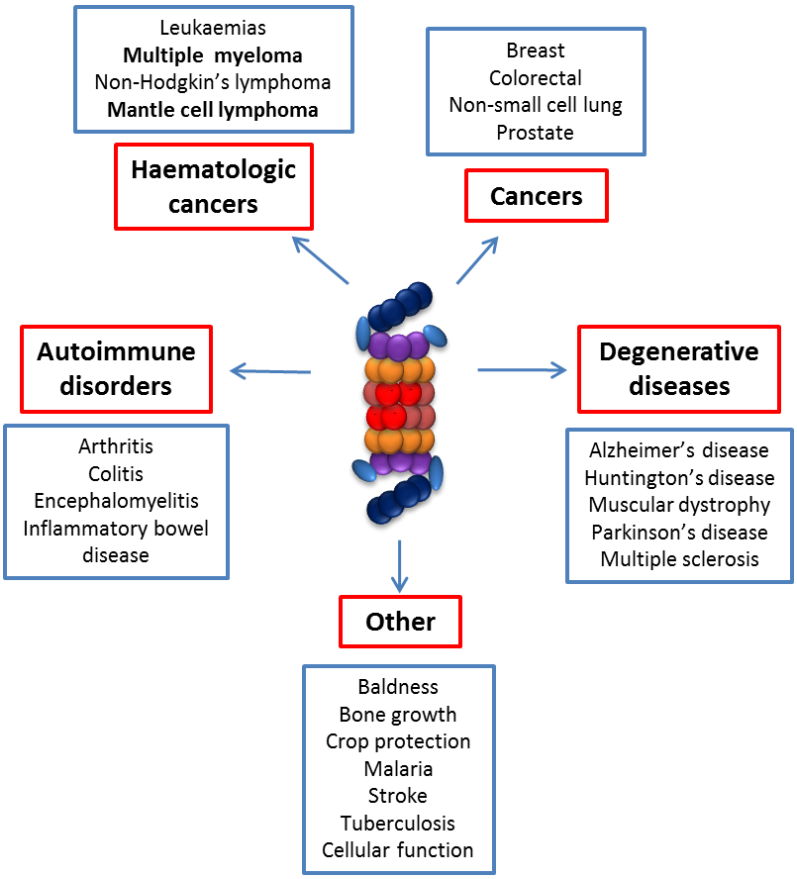


Figure 10 – Therapeutic potential of proteasome inhibition - Classes and examples of diseases and disorders for which proteasome inhibitors show therapeutic potential. Bold denotes that proteasome inhibitors are currently FDA approved for a form of treatment.

Ultimately, with such diverse impacts and therapeutic potential, the proteasome is a powerful drug target. However, IP over CP as well as active site selectivity is sought to potentially increase the therapeutic index. Between more diverse variants, species specific proteasome targeting over human proteasomes, has recently been achieved in *Plasmodium falciparum* (Li *et al.*, 2016) as well as *Mycobactirium tuberculosis* (Hsu *et al.*, 2017), through structural studies. In addition, despite few conserved changes, it has also been shown that human IP over CP selectivity is possible, although further investigation is required to bolster understanding on how to best achieve this (Kuhn *et al.*, 2009; Huber *et al.*, 2012).

1.4.4 Existing Proteasome Inhibitors

Inhibitors of the proteasome began as simple peptide aldehydes, which crucially, did not denature cells nor prevent regular functioning, or cross the blood-brain barrier (Rock *et al.*, 1994). Originally designed as analogues of chymotrypsin-like active site substrates; modifications of these peptide aldehydes soon developed more successful compounds such as MG132, a highly selective, potent, irreversible proteasome inhibitor (Lee *et al.*, 1998). This subsequently led to the discovery of bortezomib and further classes of inhibitors outlined in table 2.

Despite initial scepticism regarding toxicity and therapeutic efficacy, bortezomib gained FDA approval for MM treatment in 2003 after only phase II trials, and in 2006 for MCL treatment. By 2012, bortezomib was the first line treatment for MM, with >400,000 patients using the drug worldwide and annual sales greater than a billion pounds (Goldberg, 2012). However, common relapses and peripheral neuropathy remain issues associated with the treatment. Although bortezomib favours β5 inhibition, it does not discriminate greatly between the CP, IP or each active site and some off-target serine protease inhibition is observed (Groll *et al.*, 2006a). New classes of inhibitors are in continued development with different modes of action aiming to overcome these issues.

The second generation inhibitor carfilzomib gained FDA approval in 2012 for advanced MM (Siegel *et al.*, 2012; Herndon *et al.*, 2013) as well as relapsed and refractory MM combination therapies by 2016. This displays preference to β5 sites over β1 and β2, associated with lower toxicity; however it is not specific towards the CP or IP.

In addition, the first orally available proteasome inhibitor, ixazomib was FDA approved in MM combination therapy in 2015, predominantly targeting the β5 site (Moreau *et al.*, 2016). This acts as a prodrug shown in figure 11 that reveals the same boronate warhead group as bortezomib, but has a quicker dissociation of reversible binding.

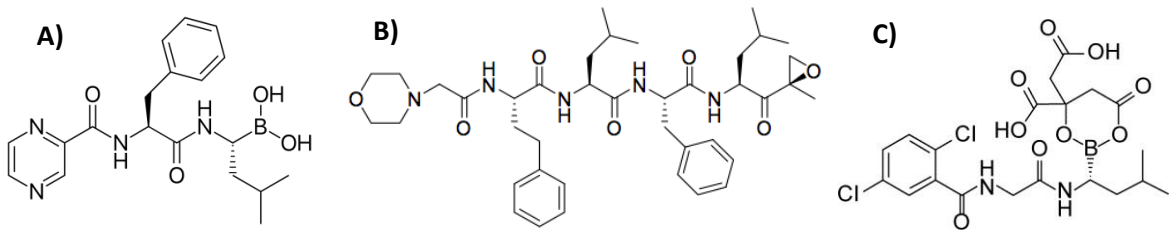


Figure 11 - Structures of FDA approved proteasome inhbitors – A) Bortezomib, B) Carfilzomib, C) Ixazomib.

Of existing proteasome inhibitors, the β5 site is most commonly blocked, although some also co-target other sites (Beck *et al.*, 2015). It has been shown that sensitivity to proteasome inhibitors can be enhanced by β1 or β2 selective inhibitors, to a greater extent than existing combination therapies (Geurink *et al.*, 2013). This can be explained by resistance mechanisms such as mutations within the active site reducing binding affinity of inhibitors or overexpression of the effected subunits (Manansanch and Orlowski, 2017).

Novel proteasome inhibitors are often created by altered peptide backbones to improve the compound's interactions with the proteasome active sites (Micale *et al.*, 2014). This is evident with carfilzomib as displayed in figure 11 where bulky and characteristic amino acid R-groups complements the specificity pockets of the active site. In some cases, non-natural amino acids can improve the potency (Geurink *et al.*, 2013) and influence the pharmacokinetic drug properties. Since hydrophobic residues are more permeable to the cells, β5 selective drugs have shown greater success rate in development (Kisselev *et al.*, 2006). Clinical trials are ongoing for oprozomib targeting β5 in haematological malignancies and marizomib targeting all active sites in a variety of cancers (Manansanch and Orlowski, 2017).

As well as targeting the catalytic centre, it is worth mentioning the potential of alternative sites of inhibitor binding. Drugs focussed on targeting active site binding pockets and non-catalytic residues have shown pharmacokinetic benefits in comparison to directly blocking Thr1 (Beck *et al.*, 2015; Duibella *et al.*, 2015). In addition, the 19S subunits contain multiple targets for drug interaction that can impact the recognition and acceptance of proteins for degradation (Kisselev *et al.*, 2012). In addition, the molecular pathways involved in the formation of the CP and IP are potential targets undergoing investigation.

The success of existing proteasome inhibitors validates the potential of the proteasome as a drug target, although further investigations are required to improve on current compounds. A key strategy to ameliorate side effects is to target only the proteasome over other proteases and it has also been shown that cytotoxicity in MM cell lines correlates with loss of active site specificity (Gräwert *et al.*, 2012; Geurink *et al.*, 2013). Therefore IP and active site selective compounds are sought after.

1.4.5 Immunoproteasome inhibitors

Given the proposed therapeutic benefits of IP selective inhibitors, discussed in 1.4.3 Therapeutic potential, this has become a focus for many leading research groups. Most existing proteasome inhibitors block the active sites of both the CP and IP at similar potencies. However, a significant recent advancement within the field has been the identification of structural differences in the binding pockets of CP versus IP (Huber *et al.*, 2012). This has led to structure-based design of IP inhibitors that show a greater affinity to the substituted residues within IP active sites over corresponding CP sites. Basic, characteristic changes are summarised in figure 6.

Carfilzomib inhibits both β5c and β5i at similar potencies although interestingly, further studies with β5c knockout in lymphoma cell lines retained a similar potency; suggesting therapeutic effects of β5i inhibition alone (Singh *et al.*, 2010; Miller *et al.*, 2013). From this, investigations into epoxyketone derivatives produced PR-924 with up to 250-fold selectivity towards β5i over β5c (Huber *et al.*, 2016; Parlati *et al.*, 2016). With this available, testing the β5i selective PR-924 on normal cells did not induce apoptosis suggesting a method for targeting only diseased cells (Singh *et al.*, 2010). Ho *et al.* (2007) and Kuhn *et al.*, (2009) also researched synthetic compounds to selectively inhibit β1i and showed sufficient potency to trigger cell death with *in vivo* mouse and *in vitro* hematologic malignancies. However in contrast, Parlati *et al.* (2009) showed reduced effect of IP selective compounds alone on malignant cells. Key pathways such as those outlined in figure 9 were not triggered, compared to use in combination with β5c inhibitors or inhibitors targeting all active sites. PR-957 is another epoxyketone with up to 20-fold greater inhibition at β5i over β5c (Huber *et al.*, 2016).

Whilst most inhibitors in development continue to target β5, there are β1 (Ho *et al.*, 2007; de Bruin *et al.*, 2014) and β2 (Geurink *et al.*, 2013; Koroleva *et al.*, 2015) selective inhibitors showing that active site specificity is possible despite few structural changes. Although there is also the added complexity from allosteric effects on β5 sites occasionally observed (Ho *et al.*, 2007).

From PR-957 and PR-924, research groups have successfully improved potency and selectivity of new, synthetic compounds through structure-based design (de Bruin *et al.*, 2014; Duibella *et al.*, 2014). Furthermore the successful approaches have been further applied to different compound backbones and warhead groups (Fan *et al.*, 2014). This includes non-peptide (Groll *et al.*, 2015; Sosič *et al.*, 2016) and non-covalent (Koroleva *et al.*, 2015) compounds exhibiting IP selective inhibition. New avenues targeting away from the active site are also being investigated, revealing influential roles of previously overlooked residues. For example, a conserved, non-catalytic Cys48

in β5i has been successfully targeted with altered inhibitor P3 groups combined with ligand stabilisation at S3 pockets (Duibella *et al.*, 2015).

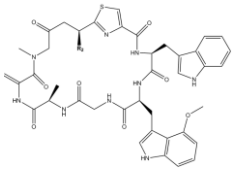
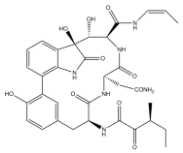
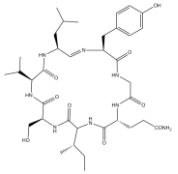
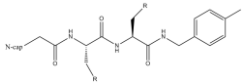
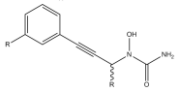
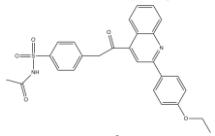
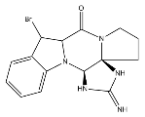
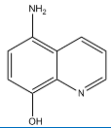
Overall, there is an expanding wealth of knowledge, with proven success that is facilitating the design of IP and active site selective compounds in a diverse range of inhibitor classes. Many resulting compounds have further supported the association of improved therapeutic index on account of targeting IP diseased state cells. As previously discussed in section 1.4.1 Classes of proteasome inhibitors, non-covalent compounds show many favourable properties as proteasome inhibitors although, there is currently little research on the area (Bellavista *et al.*, 2013).

1.5 Non-covalent proteasome inhibitors

1.5.1 Existing non-covalent proteasome inhibitors

Recent research has been orientated towards development of non-covalent inhibitors lacking a reactive group that instead display weaker intermolecular interactions and reversible inhibition. Intermolecular interactions, particularly hydrogen bonding, allow for the inhibitor to stabilise within the active site. Whilst specificity pockets allow selective interaction for the proteasome over proteases, as well as particular active sites (Lee and Goldberg, 1998). The catalytic Thr1 is not necessarily directly blocked; instead reducing the ability of the substrate to bind achieves proteasome inhibition. Known examples are summarised in table 3.

Table 3 - Current non-covalent proteasome inhibitiors – Details of class, general structure, bonding interactions and origin of example non-covalent inhibitors.

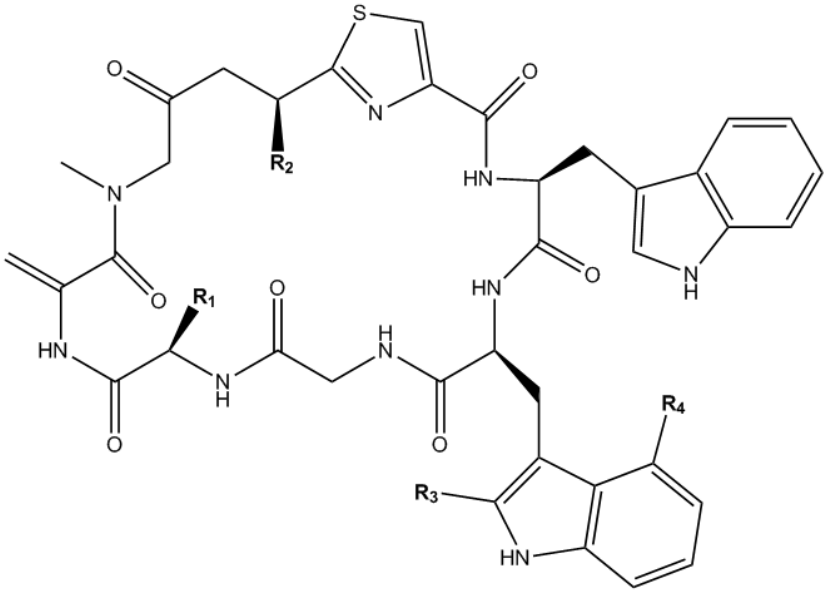
Class	Structure	Non-covalent Bonding	Compounds	Origin
Cyclic peptides		Hydrogen bonds and Van der Waals interactions at S1-4 pockets surrounding Thr1	Argyrin A ($R_2 = CH_3$) (Nickeleit *et al.*, 2008) Argyrin F ($R_2 = CH_2OH$)(Bülow *et al.*, 2010)	Naturally derived from myxobacterium
		Hydrogen bonds and Van der Waals interactions at S1-4 pockets surrounding Thr1	TMC-95A (Groll *et al.*, 2006c)	Naturally derived from apiospora
		Hydrophobic interactions at S1 pocket	Scytonemide A&B (Krunic *et al.*, 2010)	Naturally derived from cyanobacterium
Non-cyclic peptides		Occupy S1-4 pockets	C- and N- terminally capped dipeptides (Blackburn *et al.*, 2010)	Synthetic
Non-peptide inhibitors		Occupies S1 and S3 pockets, not Thr1 directly	Modified hydroxyurea (Gallastegui *et al.*, 2012)	Synthetic
		Van der Waals, hydrophobic and hydrophilic interactions. Single polar interaction.	Sulphonamide Compound 4 (Beck *et al.*, 2015)	Synthetic
		Targeting only S3 pocket of B5 site	Compound 4 indolo analogue of dibromophakellin (Beck *et al.*, 2015)	Natural – based on pentacyclic alkaloid
Allosteric inhibitors	Proline-arginine rich peptide	Non-competitive binding to α7 subunit	PR-39, PR-11 (Gaczynska *et al.*, 2003)	Naturally derived from porcine bone marrow
		Non-competitive binding to α7 subunit, possible metabolite binding β-sites simultaneously	5-amino-8-hydroxyquinoline (Li *et al.*, 2010)	Synthetic

TMC-95A is a naturally derived, cyclic tripeptide that inhibits all CP active sites and shows a slight preference towards β5c. The inhibitor does not undergo any conformational changes when non-covalently binding, providing favourable entropy. Despite displaying lower potency, structural based optimisation can be used to enhance binding affinity in synthetic analogues, as demonstrated by (Groll *et al.*, 20006c, 2010). Argyrin A is also a cyclic peptide, although formed of 8 amino acids. These display well characterised anti-tumour effects through binding all active sites of the CP (Nickeleit *et al.*, 2008, Loizidou and Zeinalipour-Yazdi, 2014). In addition, smaller, non-cyclic peptides have shown strong inhibition. From high throughput screening, (Blackburn *et al.*, 2010) identified a C- and N- terminally capped tripeptide, subsequently optimised to develop potent, non-covalent, di-peptide proteasome inhibitors as shown through x-ray crystallography. Other non-covalent inhibitors show non-competitive modes of action, in some cases binding far from the active site on alpha subunits proposed to cause conformational changes and disrupt 19S interaction (Gaczynska *et al.*, 2003, Li *et al.*, 2010).

Overall, a diverse class of inhibitors vary in binding mode and origin, with a mixture of peptide, non-peptide, natural and synthetic compounds showing non-covalent, yet potent inhibition. Understanding the energetically favourable interactions and orientations associated with specificity will help to facilitate development of future inhibitors.

1.5.2 The argyrin family

Naturally derived products have been successful contributors to cancer therapeutics, including proteasome inhibitors, with several outlined in table 2-3. Furthermore, although many compounds have been synthetically optimised, these originated from natural sources such as microbial metabolites (Gräwert *et al.*, 2012). Recently, the argyrin A-F cyclic peptides, derived from the myxobacterium *Archangium gephyr,* have emerged as an exciting family of compounds (Sasse *et al.*, 2002). Subtle differences in structure between analogues are shown in figure 12. These cyclic peptides have been identified as proteasome inhibitors that target all 3 active sites, with varying potencies (Bülow *et al.*, 2010).



Argyrin	R_1	R_2	R_3	R_4
A	H	CH_3	H	OCH_3
B	CH_3	CH_3	H	OCH_3
C	H	CH_3	CH_3	OCH_3
D	CH_3	CH_3	CH_3	OCH_3
E	H	CH_3	H	H
F	H	CH_2OH	H	OCH_3
G	CH_3	CH_2OH	H	OCH_3
H	H	H	H	OCH_3

Figure 12 - Structure of argyrin analogues A-H – Table represents R group variants amongst different analogues (Adapted from: Vollebrecht *et al.*, 2002).

Interestingly, the anti-tumour activity of argyrin A and F is non-covalent CP inhibition dependent on $p27^{kip1}$ expression. This is distinct from existing bortezomib treatment that acts through inhibitory kappa B alpha (IκBα) stabilisation. Therefore may offer a novel molecular pathway effect that can help overcome resistance. Furthermore, the reduced vascular endothelial growth factor (VEGF) expression from argyrin A treatment has anti-angiogenic properties leading to loss of endothelial cell adherence, to target tumour vascularisation (Nickeleit *et al.*, 2008). These effects were shown in purified proteasomes, multiple cell lines and xenograft tumours. Gene expression profiling in MCF7 cells revealed argyrin A treatment to affect approximately 500 genes, all related to the proteasome. In comparison, bortezomib treatment resulted in 10900 genes with altered expressions, including effects beyond the proteasome (Nickeleit *et al.*, 2008). The targeted action of argyrin A is therefore promising for drug development with lower cytotoxic effects. In

addition, a different mode of action can overcome relapsed and refractory cancer patients where resistance has been developed.

Following these findings, argyrin F was identified to have even more pronounced anti-tumoural activities, in the same manner as argyrin A (Bülow *et al.*, 2010). Recent studies show argyrin F as a potential combination therapy treatment for pancreatic adenocaracinoma. In vitro human cell line studies and in vivo mouse models revealed reduced cell proliferation due to dose and time dependant upregulation of tumour suppressor cyclin kinase inhibitors $p27^{kip1}$ and $p21^{waf1/cip1}$, alongside a reduction in cell migration, invasion, angiogenesis and tumour growth (Chen *et al.*, 2017).

The ability of argyrins A and F to exhibit biological activity at low concentrations, with targeted actions, distinct molecular pathways and proven anti-cancer effects *in vitro* and *in vivo*; make these naturally derived analogues promising therapeutics. To date, remaining analogues have not yet received further investigation as proteasome inhibitors and none have been tested with the IP.

1.5.3 Argyrin B

Differing from argyrin A, analogue B possesses an α-amino butyric acid in place of alanine, creating an additional methyl group extension. The same is observed in comparison to argyrin F, although this also contains an additional hydroxyl group at R2 (Vollbrecht *et al.*, 2002), as shown in figure 12. With similar structures it is therefore likely that argyrin B exhibits closely related mechanisms to its analogues.

Argyrin B was initially discovered in screening for antibiotics and has displayed some antibacterial as well as antifungal activities. In human B cells, argyrin B exhibits immunoglobulin G (IgG) inhibition which combined with murine studies showing reduced activity of T and B lymphocytes, shows strong immunosuppressive effects of the compound (Sasse *et al.*, 2002). This area has received greater attention recently with further antibacterial activity and associated mechanisms outlined (Jones *et al.*, 2017). Despite the promising actions of analogues A and F as proteasome inhibitors, there is little research on argyrin B in this role. Only basic percentage remaining activity from purified CP and colon cancer (SW-480, HCT116) cytotoxicity assay data is known (Bülow *et al.*, 2010).

With few cyclic peptides and non-covalent inhibitors researched as proteasome inhibitors as well as no current investigations into argyrin action at the IP, there remains a gap in current research.

Argyrin B is a promising compound for further investigation as a CP and IP inhibitor. Furthermore, with crystal structures available there is the opportunity for computational structural studies such as molecular docking.

1.6 Approaches for investigating proteasome inhibitors

1.6.1 Laboratory based inhibition studies

Proteasome activity assays are important tests for the efficacy and specificity of compound inhibition. These can be performed as purified assays or in cell lines and there are a wide range of fluorogenic substrates available that are designed to be specific at particular active sites. Typically substrates are short peptides attached to a fluorescent component that becomes cleaved upon binding the catalytic site (Kisselev and Goldberg, 2005). The release of free fluorogenic substrate, detectable at particular wavelengths, is then proportional to the activity of the enzyme. To achieve an overall view of inhibition it is required to measure activity at each active site. These assays can be performed with commercially available purified 20S CP or IP, and is a common approach used, as outlined in table 22. This allows a variety of kinetic tests that can establish inhibitor potency and actions. Tests in cell lines with fluorogenic substrates incur additional challenges to allow the substrate access to the proteasome (Harris *et al.*, 2001). Commonly cells are lysed and the 26S proteasome is broken down, losing the 19S cap. Therefore a complete picture of cellular interactions is difficult to model with fluorogenic substrates.

To overcome this, it is possible to use chemical proteasome probes to measure activity in live cells or their lysates (Berkers *et al.*, 2005). Probes containing a potent inhibitor are able to covalently bind Thr1. The attached motif can then be detected using antibodies in enzyme linked immuno sorbent assay (ELISA) methods (Kuhn *et al.*, 2009). Alternatively fluorescent motifs can be visualised or analysed with flow cytometry, allowing tests at different stages of the cell cycle (Muchamuel *et al.*, 2009). Using cell based assays it is also possible to measure proteasome activity *in vivo*. Proteasome substrates with a ubiquitin and fluorescent tag, such as green fluorescent protein (GFP) or luciferase, can enter the cell and undergo degradation, thereby reducing fluorescence (Dantuma *et al.*, 2000). Upon inhibition of the proteasome, the GFP and luciferase remain present and produce intracellular fluorescence. However, using this method is difficult to account for possible degradation through other mechanisms.

With established inhibition, crystallisation of compound and proteasome complexes can reveal details of the binding mode. This will help to elucidate specific residues that the inhibitor binds or interacts with and plays a key role in facilitating understanding of inhibitor selectivity. As well as laboratory methods, computational modelling can be used to predict binding of the inhibitor and screen libraries of compounds.

1.6.2 Computational methods

The CP can be purified from erythrocytes, and crystal structures have recently been elucidated in humans (Harshbarger *et al.*, 2015, Huang *et al.*, 2016). However, difficulties in isolating IP due to their induced formation means that human IP crystal structures have not yet been produced with only yeast and mice IP crystal structures available (Huber *et al.*, 2012). Homology models can be used, for example humanised IP structures through sequence substitution of human amino acid sequences with existing yeast structural data. This approach has been successfully used to model drug interactions at human IP active sites (Bellavista *et al.*, 2013, Loizidou and Zeinalipour-Yazdi, 2014).

With structural data of the enzyme and active site, a variety of computational software can be used to model proteasome-ligand interactions. Although not as accurate as generating crystallisation data with the enzyme and compound, molecular docking can be a useful tool for indications of binding modes and interactions.

Molecular modelling approaches simulate the enzyme and ligand binding; predicting the most energetically favourable pose of interaction. Various algorithms can predict binding energy, thereby affinity for multiple conformations and orientations on the target. Most docking tools are able to identify non-covalent interactions, including hydrogen bonding, Van der Waals forces and electrostatic interactions, with associated scoring functions. Whilst some are also capable of covalent docking where the ligand and target undergo a reaction to establish a bond (Jones *et al.*, 1997).

Overall, molecular docking has been proven as a valuable resource in drug design and development, including applications to inhibitors of the proteasome. With human CP and murine IP structures now available, it is becoming an increasingly used research tool. However, it is important to note that scoring functions have proven inconsistent between different software

and in comparison to laboratory studies (Chen, 2015). It is therefore preferable that molecular docking is used to supplement laboratory research.

1.7 Aims

This project investigates the inhibitory effect and specific interactions of argyrin B at each active site of the CP and IP, with the aim of identifying factors that confer IP over CP selectivity. This can be achieved through the following sub aims:

- Establish argyrin B inhibitory activity of the CP and IP, testing at β1 and β5 active sites using *in vitro* assays of purified 20S proteasome.

- Determine argyrin B IC_{50} and inhibition constants at β1 and β5 active sites of the CP and IP.

- Develop a suitable homology model of human IP active sites for use in molecular docking.

- Predict molecular interactions of argyrin B at each active site of the CP and IP, through molecular docking simulations.

Chapter 2. Methods

2.1 Laboratory methods

2.1.1 Materials

The following reagents and materials outlined in table 4-5 were used for enzymatic assays.

Table 4 - Proteasome and immunoproteasome buffer contents and concentrations.

Name	Constituents	Concentration
Proteasome assay buffer, pH 7.5 (BML-KI340-0020, Enzo Life Sciences, Exeter, UK)	Tris/HCl	50 mM
	KCl	25 mM
	NaCl	10 mM
	$MgCl_2$	1 mM
	SDS	0.03% (w/v)
20S proteasome (purified human erythrocyte) enzyme, 1 mg/ml in buffer (see right), pH 7.2 (BML-PW8720-0200, Enzo Life Sciences, Exeter, UK)	Tris/HCl	20 mM
	EDTA	1 mM
	DTT	1 mM
	Sodium azide	1 mM
20S immunoproteasome (purified human) enzyme, 0.5 mg/ml in TEAD buffer (see right), pH 7.4 (BML-PW9645, Enzo Life Sciences, Exeter, UK)	Tris/HCl	20 mM
	DTT	1 mM
	Sodium azide	1 mM

Table 5 - Laboratory reagents and materials.

Consumable name	Details	Product number
96 well microplates	½ volume NBS, white	80-2406, Enzo Life Sciences, Exeter, UK
AMC Standard	7-amino-4-methylcoumarin	BML-KI122-0014, Enzo Life Sciences, Exeter, UK
Argyrin B	$C_{41}H_{46}N_{10}O_8S$, MW 838.9Da	Donation from Novartis, Surrey, UK
β1c substrate	Z-Leu-Leu-Glu-AMC	BML-ZW9345, Enzo Life Sciences, Exeter, UK
β1i substrate	Ac-Pro-Ala-Leu-AMC	S-310, BioTechne, Abingdon, UK
β5c substrate	Suc-Leu-Leu-Val-Tyr-AMC	BML-P802-9090, Enzo Life Sciences, Exeter, UK
β5i substrate	Suc-Leu-Leu-Val-Tyr-AMC	BML-P802-9090, Enzo Life Sciences, Exeter, UK
DMSO	Dimethyl sulphoxide	BP231-1, Fisher Scientific, Loughborough, UK
Epoxomycin	$C_{28}H_{50}N_4O_7$, MW 554.7g/mol	BML-PI127-9090, Enzo Life Sciences, Exeter, UK
Methanol	MeOH, HPLC grade 99%	268280025, Acros Organics, Loughborough, UK
SDS	Sodium dodecyl sulphate	BML-KI341-0012, Enzo Life Sciences, Exeter, UK
Vinyl sulfone	Ada-$(Ahx)_3$-$(Leu)_3$-vinyl sulfone	BML-AW9155, Enzo Life Sciences, Exeter, UK

Substrate and inhibitor reagents were dissolved in dimethyl sulphoxide (DMSO) for stock solutions and subsequently diluted in proteasome assay buffer to desired concentrations. Although required for compound solubility, DMSO levels were minimised to reduce any potential toxic

effects. Specific concentrations used in purified assays are outlined in supplementary table 1. Assay reagents were added to 96-well plates to a final volume of 50 µl per well, throughout. Concentrations of CP and IP were maintained constant at 0.1 µg/well, whilst inhibitor and substrate were varied according to kinetic parameters tested. A 7-amino-4-methylcoumarin (AMC) standard was prepared from a 1:2 serial dilution of 8 µM to 0.25 µM and blank.

2.1.2 Purified enzymatic assays

Purified assay positive controls were performed through the reaction of CP or IP enzyme with active site specific substrates: Z-LLE-AMC (β1c), Ac-PAL-AMC (β1i) and Suc-LLVY-AMC (β5c/i). The liberation of AMC was measured over time using a BMG Labtech fluorescence plate reader set at 355/460 nm (excitation/emission).

For inhibitor analysis, argyrin B was incubated with the enzyme at 37 °C, for 10 minutes prior to substrate addition and maintained at this temperature throughout the reaction. A range of inhibitor and substrate concentrations were used to determine K_m, IC_{50} and K_i values. Negative controls performed for each assay were low concentrations of epoxomycin and vinyl sulfone; potent inhibitors of β5c/i and β1c/i active sites, respectively. Blanks were performed with substrate only and additional controls for solvent (DMSO) concentration were used where applicable.

Tests were run for at least one hour and 50 reading cycles. An appropriate gain was set according to background fluorescence and the speed of reaction, to ensure the equipment's detection limit of 250,000 was not met within an hour of running. A summary of the assay process, including plate reader settings are provided in figure 13.

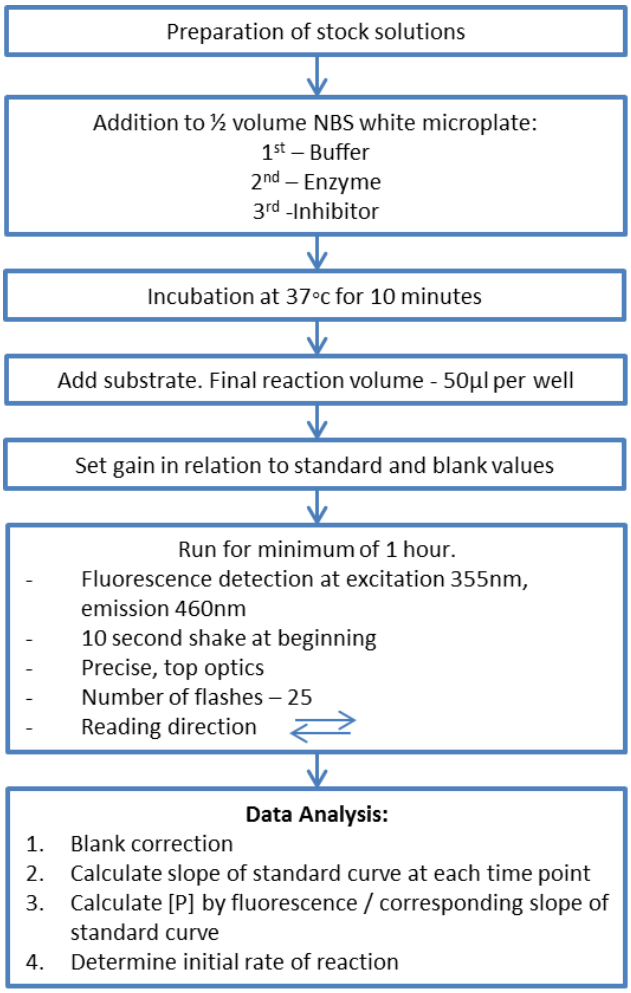


Figure 13 – Purified kinetic assay method flowchart - Using 20S proteasome enzyme, subunit specific substrates (Z-LLE-AMC, Ac-PAL-AMC & Suc-LLVY), argyrin B inhibitor and controls of epoxomycin or vinyl sulfone.

Using 0.1 µg/well of enzyme, a range of at least 7 substrate concentrations each in triplicate, were used to generate Michaelis-Menten plots for Michealis-Menten constant (K_m) analysis at each active site. Subsequent K_m values were used as the single substrate concentration in IC_{50} plots that covered a logarithmic range of at least 10 argyrin B concentrations. 3 independent repeats were performed for each IC_{50} active site assay. For kinetic assays to determine inhibition constant (K_i), 4 concentrations of inhibitor were used, each at 5 different substrate concentrations. Argyrin B concentrations ranged from estimated IC_{50} value (β1c = 183.7 µM, β1i = 10.4 µM, β5c = 11.4 µM, β5i = 10.3 µM) x 0, 0.33, 1 and 3, whilst substrate concentration covered K_m (β1c = 95.4 µM, β1i = 69.9 µM, β5c = 72.4 µM, β5i = 89.8 µM) x 2.5, 1.25, 0.625, 0.3125 and 0.15625.

2.1.3 Data analysis and statistics

Each test was performed with triplicates at every control and concentration variant, whilst AMC standards were in duplicate. Firstly, values for substrate only blanks were subtracted from corresponding substrate concentration data. An AMC standard curve was produced to convert all data from fluorescence, into product concentration. DMSO correction was performed where equivalent concentrations to those from argyrin preparation elicited an effect compared to control. For each replicate, the initial rate of reaction was determined from the linear phase of the graph at which less than 10% of substrate had been consumed.

Using GraphPad Prism 6 software, non-linear curve fitting analysis was used for calculation of K_m, IC_{50} and K_i values. The preferred model to determine K_m was a Michaelis-Menten plot, based on equation 1 where 'y' represents initial velocity and 'x' refers to substrate concentration. Graphs are depicted with average values, displaying the standard error of mean (SEM), calculated from 3 replicates. To enhance statistical accuracy of all non-linear analysis, the degrees of freedom were increased by treating all replicates as individual points rather than a single average value.

Equation 1 – Michaelis-Menten model – Equation fit used in GraphPad analysis.

$$Y = \frac{Vmax * X}{(Km + X)}$$

For IC_{50} analysis, velocities were expressed as a percentage of control values with an expected range from approximately 100% to 0%. As such, a normalised response curve was fit against logarithmic inhibitor concentration. A variable slope function was allowed, fitting a hill slope determined from the data with analysis based on equation 2. Data were reported with mean, SEM and 95% confidence interval. 95% confidence was determined as the mean plus and minus 2.306 times the standard error of mean; based upon the 2-tailed inverse student's t-distribution test with 9 degrees of freedom. Graphpad analysis or Excel syntax '=T.INV.2T(0.05,N-1)' was used for any variation in degrees of freedom.

Equation 2 – IC_{50} analysis equation fit - Model for log(inhibitor) vs normalised response - variable slope.

$$Y = \frac{100}{(1 + 10^{((LogIC50 - X) * HillSlope))})}$$

A standard, unpaired, t-test shown in equation 3 (Spence, 1976), analysed significance between 3 IC_{50} value repeats of argyrin B at different active sites. A 95% confidence level was estimated for the difference between the means and assigned a probability value, deemed significant if $p<0.05$.

Equation 3 - Standard t-test between IC_{50} values – Used between different active sites and CP vs IP analysis. Calculated from 3 independent repeats with associated standard error.

$$t = \frac{IC50_A - IC50_B}{\sqrt{SE_A{}^2 + SE_B{}^2}}$$

To determine K_i values, tests were first performed to determine the best-fit model between: competitive, non-competitive, uncompetitive and mixed inhibition. To distinguish between competitive and other models, the graphical representation alongside K_m and V_{max} values were used. Additionally, for a more quantifiable measure, Akaike's information criterion (AICc) was used to select the most suitable model (Motulsky and Christopolous, 2003). For each model, the AICc is calculated based upon the sum-of-squares that measures goodness of fit, and degrees of freedom. In the case of 2 nested models whereby one is a special case of the other, for example non-competitive and mixed inhibition, an extra sum of squares F test can also be used. Further to the AICc test, the F test hypothesis testing approach was used to determine statistical significance at $p<0.05$. Here, unless $p=<0.05$, the simpler model was chosen.

Once the most suitable model was defined, simultaneous non-linear regression (SNLR) analysis was performed with the relevant Michaelis-Menten derived regression analysis formulae and $1/y^2$ weighting applied. Estimates of the inhibition constant K_i were reported with SEM and 95% confidence intervals.

In addition to SNLR analysis, equivalent analysis methods were used to further validate calculations. Dixon plots of 1/v against [I], combined with Cornish-bowden plots of [S]/v against [I], allowed for determination of inhibition modality as well as a K_i estimation. In addition, the y-intercepts from Woolf plots of [S]/v against [S], produced values for K_m/V_{app}. An additional plot of K_m/V_{app} against [I] subsequently yields a K_i value from the x-intercept. Analysis was performed using Microsoft Excel linear regression and errors from 'LINEST' function.

2.2 Computational methods

2.2.1 Software

The computational software and websites used have been summarised in table 6.

Table 6 - Computer software and websites.

Software	Source
AutoDock (v4.2.6)	The Scripps Research Institute (California, USA) (Morris *et al.*, 2009) http://autodock.scripps.edu
AutoDockTools (v1.5.6)	The Scripps Research Institute (La Jolla, USA) http://mgltools.scripps.edu
Avogadro (v.1.2)	Avogadro Chemistry (Hanwell *et al.*, 2012) http://avogadro.openmolecules.net
ChemDraw (v14.0)	Perkin Elmer Informatics (Cambridge, US)
EMBL-EBI Clustal Omega	European Bioinformatics Institute (Hinxton, UK) (McWilliam *et al.*, 2013) http://www.ebi.ac.uk/Tools/msa/clustalo/
GraphPad Prism 6	GraphPad Software (La Jolla, USA) (GraphPad 6.04, 2014) www.graphpad.com
Minitab 17	Minitab Ltd (Coventry, UK) (Minitab 17.1.0, 2013)
Open Eye Scientific Software-OEDocking FRED (v3.2.0.2)	http://www.eyesopen.com (Santa Fe, NM) (McGann, 2012)
PyMol (v1.7.4.5)	Schrodinger LLC (Delano *et al.*, 2002) https://www.pymol.org
RSCB Protein Data Bank	Research Collaboratory for Structural Bioinformatics (Berman *et al.*, 2000) http://www.rcsb.org
SwissPDB viewer (V.4.10)	Swiss Institute of Bioinformatics (Guex and Peitsch, 1997) http://spdbv.vital-it.ch/
UniProt	UniProt consortium (2014) (UK, CH, USA) http://www.uniprot.org/

2.2.2 Structural data preparation

All 3-dimensional structures were obtained from open access to RCSB Protein Data Bank (PDB) and prepared using the molecular graphics package PyMOL (v1.7.4.5). An argyrin B structure was isolated from PDB:4FN5 (Nyfeler *et al.*, 2012) and energetically optimised to a stable structure using Avogadro (Hanwell *et al.*, 2012). Bonds connecting the tryptophans to the cyclic ring, as well as the Trp2-OMethoxy bond were allowed rotational freedom, as shown in figure 14.

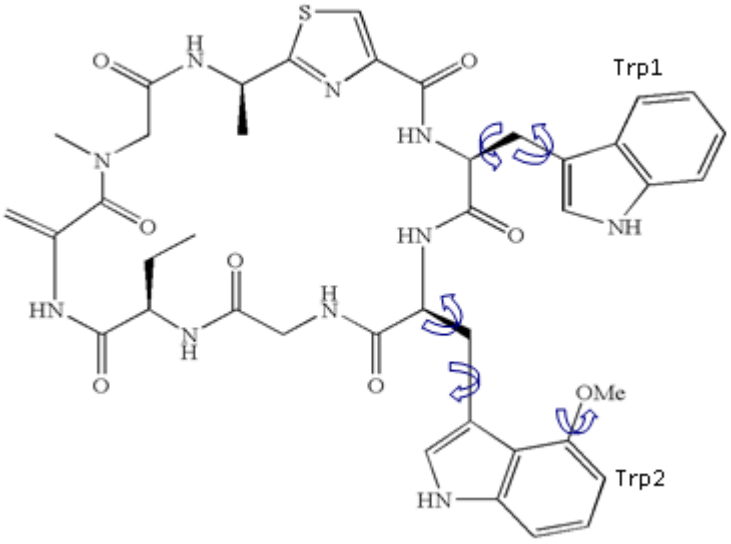


Figure 14 – Chemical structure of argyrin B - Displaying rotatable bond positions when used in AutoDock modelling simulations.

The Human constitutive 20S proteasome structural data was available from PDB:4R3O (Harshbarger *et al.*, 2015). All active sites were cut for residues within 28 Å from the catalytic Thr1 position of the active subunit chain which includes overlap into neighbouring chains, as shown in figure 15. β1, β2 and β5 cuts were all taken from the same β-ring on the same monomer. Any charged iodide, potassium and chloride ions as well as waters were removed from the structures for compliance with the modelling software.

IP structural data was obtained from PDB:3UNH (Huber *et al.*, 2012) murine IP and subsequently modified in order to create a humanised IP model. Active site cuts were initially prepared as previously described for the CP. Human active site IP sequences β1i, β2i and β5i (UniProt: P28065, P40306, P28062, respectively) were aligned to murine IP FASTA sequences using EMBL-EBI ClustalOmega, EMBOSS Smith-Waterman alignment algorithm. Utilising the amino acid mutation utility of SwissPDB viewer (v.4.10) individual amino acid substitutions were performed from murine IP structural data to corresponding human IP active subunit sequences.

Further sequence similarity comparisons were performed using basic local alignment search tool (BLAST) with the blastp algorithm and visualised alignments were produced using PyMOL functions.

2.2.3 AutoDock simulations

AutoDock (v.4.2.6) in conjunction with AutoDockTools (v.1.5.6) were used to simulate argyrin B binding (Morris *et al.*, 2009). In each case, 'PDBQT' files were created for both the ligand and macromolecule. The grid box was set at 70x, 70y, 70z dimensions, positioned at the Thr1 co-ordinates on the active subunit chain, before performing an Autogrid simulation. All co-ordinates of Thr1 catalytic centres are summarised in supplementary table 3. The large grid box size, shown in figure 15 allows the possibility of allosteric binding. The macromolecule was set with a rigid filename and genetic algorithm with 50 runs for each test as well as further parameters summarised in supplementary table 2. A Lamarckian genetic algorithm (4.2) output was set and run through AutoDock. The final file was analysed displaying the different binding energetics at the 50 different conformations. The conformation with the lowest binding energy from each test was selected for further analysis of interactions. Each active site was run for a minimum of 10 repeats with 50 conformations per test. From the transition of unbound to ligand-protein bound states, predictions of binding energy and K_i values are generated. These are based upon intermolecular interactions including: electrostatic interactions, Van der Waals forces, hydrogen bonding and desolvation energy (Morris *et al.*, 2009).

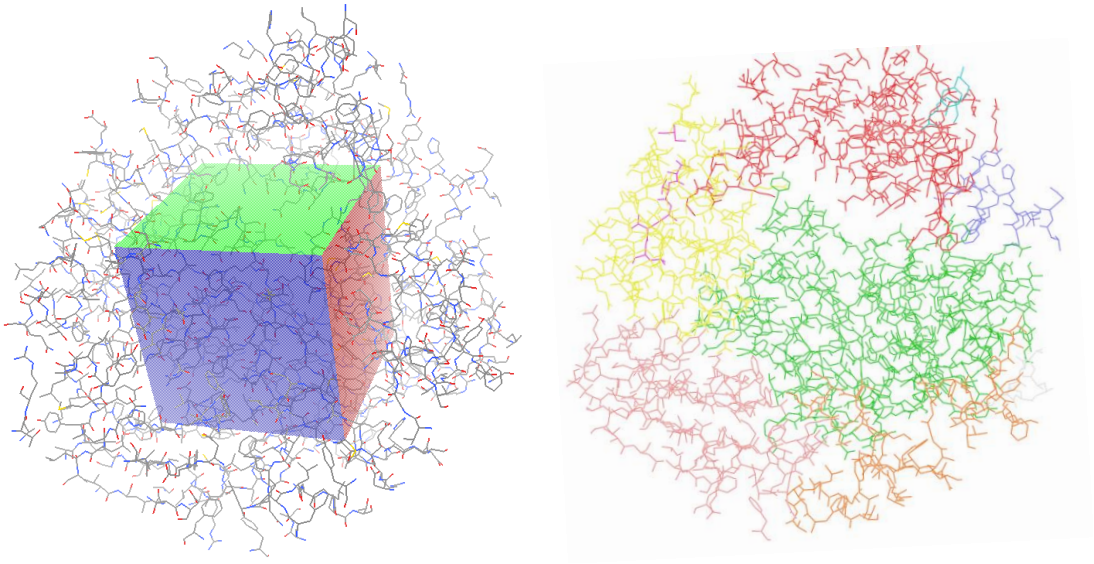


Figure 15 – Active site cut grid box size and subunit chains – Left; size of 70x-70y-70z grid box centred at Thr1 of β5c active site cut. Right; colours representing different chain active site subunits from β5c cut.

In order to validate the docking model, docking pose predictions required comparison against experimental data. To achieve this, a molecule similar to argyrin B was chosen, that also has crystallised, structural data available. PDB:1JD2 provides crystallisation studies of the non-

covalent, cyclic peptide TMC-95A in complex with the yeast 20S proteasome where binding is observed at β2 sites. Pose predictions from AutoDock simulations were compared to crystal structures, testing the frequency of similar poses. Structural preparation and docking settings followed the same aforementioned procedures with grid box coordinates shown in supplementary table 3 and 5 repeats performed of 50 runs each.

2.2.4 FRED Docking

Tests were performed to compare interactions from AutoDock simulations, run using alternative software, OEDocking FRED (v.3.2.0.2) (McGann, 2012). Identical 28 Å active site cuts to those previously used were prepared with the grid box originating from the Thr1 position at a corresponding size. Argyrin B was docked using an exhaustive search algorithm, with 100 poses using the argyrin B conformation from AutoDock simulation best-fit at each corresponding active site. Following the exhaustive search at different ligand rotations and conformations, the top scores undergo further optimised searches and are scored using Chemgauss4 that accounts for shape interactions and hydrogen bonding.

2.2.5 Statistical analysis

The binding energy estimations produced from AutoDock simulations were compared between active sites of both the CP and IP. From each set of 10 replicate best binding energies, the data were tested for normality where those with $p>0.05$ show normal distribution. For normally distributed sets, equal variance was tested between each active site. A 2-sample t-test was used for those of equal variance, whilst a non-parametric Mann-Whitney test was used for results not of equal variance, to test for significance. Generated p-values of <0.05 show statistical difference, whilst >0.05 are not statistically significant.

Chapter 3. Results

3.1 Purified enzymatic assays

3.1.1 Solvent toxicity

DMSO is required in preparation of stock solutions for solubility of compounds, as shown in supplementary table 1. As levels of DMSO remain present in the final reaction, the toxic impact of DMSO solvent alone was tested at the CP and IP.

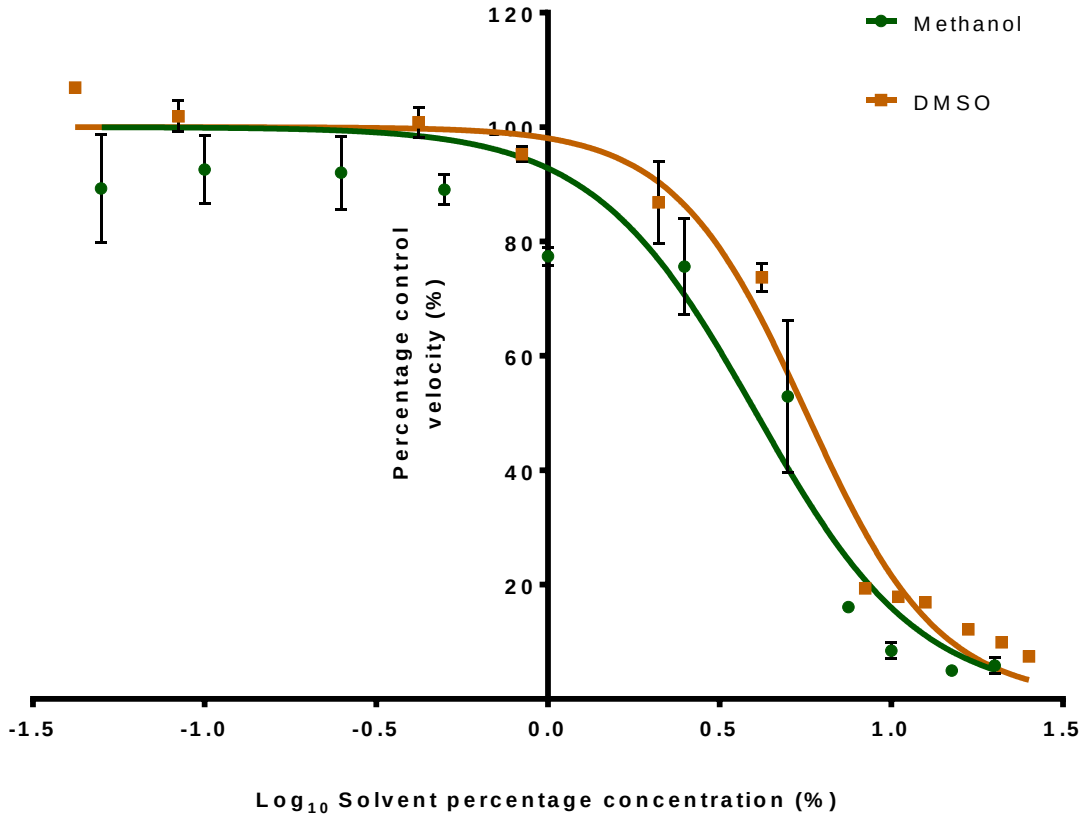


Figure 16 – Effect of solvent on CP rate of reaction – Increasing DMSO and methanol solvent concentration on a logarithmic scale, against the percentage of control CP reaction rate. β1c purified assay with Z-LLE-AMC substrate. Non-linear regression analysis fit with normalised response and variable slope. Data presented as average with standard error bars.

In addition, the impact of methanol was tested as a potential alternative solvent. Figure 16 shows a slightly lower impact of DMSO compared to methanol and no impact on CP activity at 1% or below. At higher levels, the solvents each reduced the rate of reaction significantly. However, IP tests shown in supplementary figure 1 reveal a greater sensitivity of the IP to DMSO exposure with concentrations above 0.5% showing a reduction in rate. Therefore subsequent CP tests containing a final DMSO percentage greater than 1% and IP tests with DMSO greater than 0.5% require additional DMSO control tests to account for the toxic effect of the solvent.

3.1.2 K_m values

Michaelis-Menten plots at β1c, β1i, β5c and β5i are represented in figure 17 and summarised with statistics in table 7.

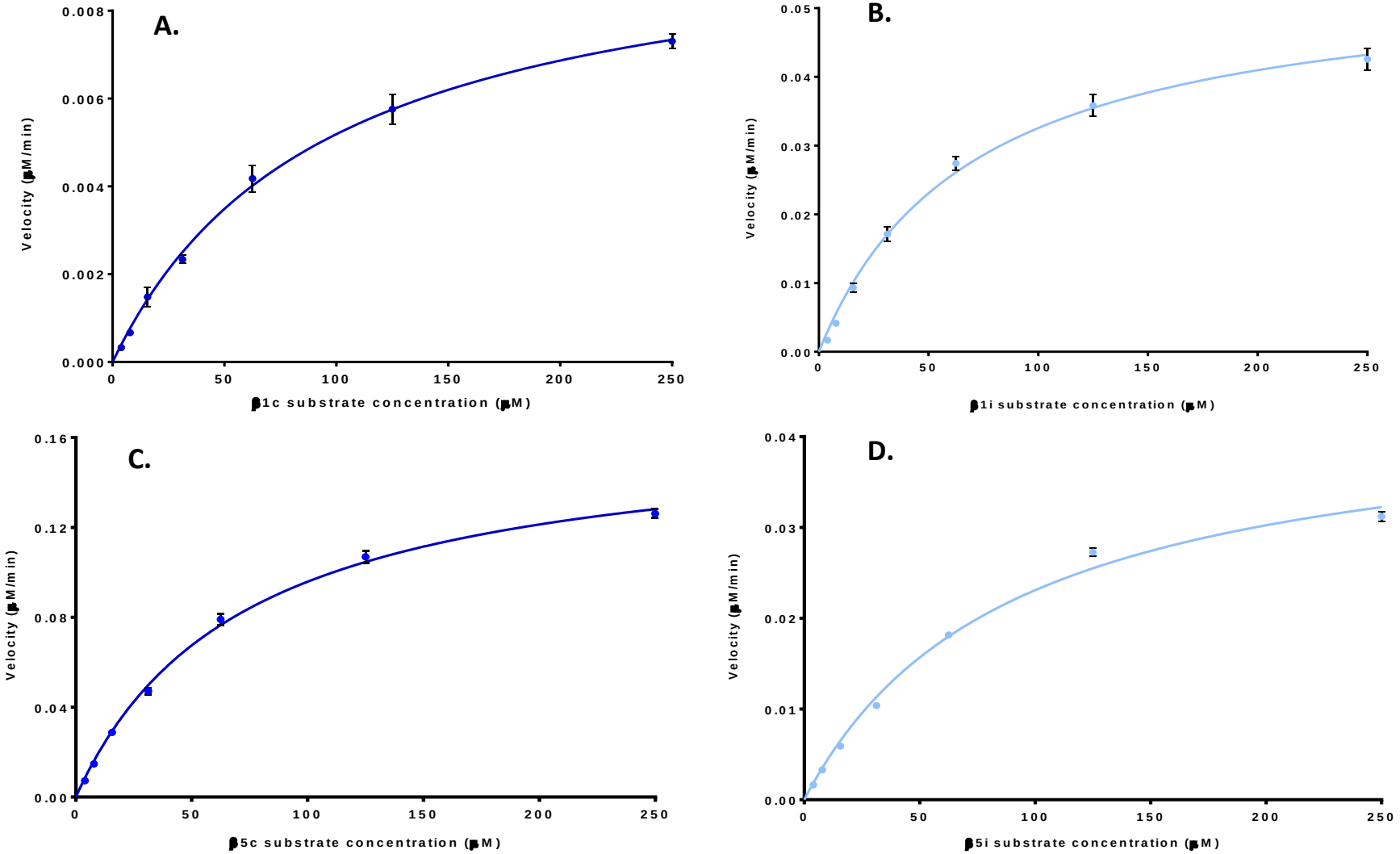


Figure 17 – Michaelis-Menten plots to determine K_m at each active site – Average velocity and SEM error bars at range of substrate concentrations, A. β1c Z-LLE-AMC, B. β1i Ac-PAL-AMC, C. β5c Suc-LLVY-AMC, D. β5i Suc-LLVY-AMC. Non-linear Michaelis-Menten analysis generated K_m values: β1c = 95.4 µM, β1i = 69.9 µM, β5c = 72.4 µM, β5i = 89.8 µM.

Table 7 - Summary of K_m values at β1 and β5 sites of CP and IP.

	β1c	β1i	β5c	β5i
K_m (μM)	**95.4**	**69.9**	**72.4**	**89.8**
SEM	9.33	6.80	4.22	7.10
95% CI	75.8 - 115.0	55.6 - 84.0	63.6 - 81.2	74.9 – 104.6
V_{max} (μM/min)	0.010	0.055	0.165	0.044
R-square	0.9876	0.9858	0.9948	0.9916

Figure 17 data shows small SEM bars and well fit analysis curves producing high r-square values. However, 95% CI levels remain broad due to low degrees of freedom.

Lineweaver-burk analysis shown in supplementary figure 3 provide an alternative method of analysis, generating K_m = 119.3 μM and 115.7 μM for β5c and β5i respectively. However, the increased influence of outer points affected the overall orientation of the regression fit, as evidenced by K_m values of 89.9 μM (β5c) and 100.1 μM (β5i) after removal of the lowest substrate concentration.

3.1.3 IC_{50} Values

For all tests shown, negative controls displayed near complete rate inhibition at low concentrations of epoxomycin or vinyl sulfone.

Figure 18 (A) analysis produces argyrin B IC_{50} values of 8.76 μM at β1i and 146.5 μM at β1c from 3 independent tests. Supplementary figure 2 shows trial, independent tests at β1i and β1c, revealing similar final values for IC_{50} at 9.6 μM and 173.9 μM respectively. However these were performed at a substrate concentration of 50 μM and a higher range of argyrin B concentrations was required. At the β5 site, figure 18 (B) also showed a greater effect on the IP with IC_{50} values of 8.30 μM for β5c and 3.54 μM for β5i. A summary of calculated values and associated statistics are summarised in table 8.

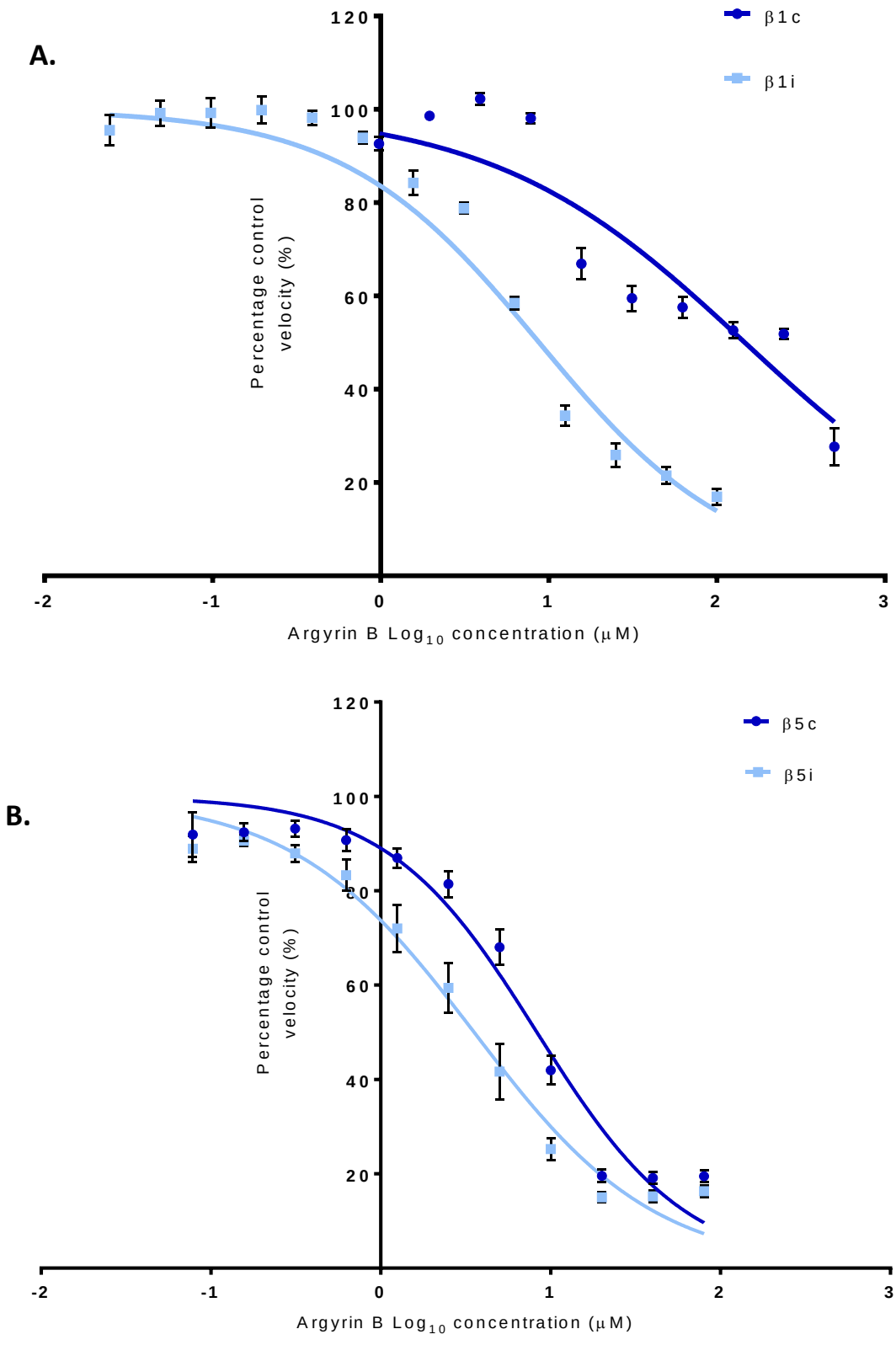


Figure 18 – Argyrin B IC_{50} plots at β1 and β5 sites of CP and IP - Logarithmic argyrin B concentration against percentage of control, initial rate velocity. Tested at [CP] and [IP] = 0.1 μg/well and [S] = K_m. Data were normalised to DMSO solvent controls where applicable. Non-linear regression analysis with variable hill slope and $1/y^2$ weighting generated IC_{50} values with respective SEM from 3 independent repeats. A) β1i IC_{50} = 8.76 μM +/-1.08, β1c IC_{50} = 146.5 μM +/-1.10. B) β5c = 8.30 μM +/- 1.07, β5i = 3.54 μM +/- 1.08.

Table 8 - Summary of IC_{50} analysis and associated statistics at β1 and β5 sites of CP and IP - * denotes statistical significance from t-test.

	β1c	β1i	β5c	β5i
Average IC_{50} (μM)	**146.5***	**8.76**	**8.30**	**3.54**
SE	1.10	1.08	1.07	1.08
95% CI	122.3 – 175.5	7.6 – 10.1	7.8 – 9.4	3.0 – 4.1
Hill Slope	-0.5784	-0.7487	-0.9881	-0.8153
R-square	0.8165	0.9271	0.9123	0.8977

Overall, β5i displayed the lowest IC_{50} value and β1c the highest. Comparing IC_{50} values from each replicate, t-tests were performed between means of each active site data with p-values detailed in supplementary table 6. At $p<0.05$, β1c IC_{50} was determined statistically significant to all other active sites whilst statistical significance was not observed between other active sites.

3.1.4 K_i values

High concentration of argyrin B at 3 x IC_{50}, 439.5 μM are required for β1c K_i range tests, resulting in high DMSO levels on account of argyrin solubility. As such, consistent with previous findings, (figure 16), high levels of DMSO alone have a significant impact on the enzyme substrate reaction (figure 19). With high DMSO levels causing a dramatic shift in reaction conditions such as proteasome concentration, the effect of argyrin B cannot be appropriately measured even with DMSO control correction. Tests at upper limits of DMSO concentrations, shown in supplementary figure 4 predict a K_i value of over 250 μM, using a single argyrin B concentration. It can be estimated that K_i is >100 μM, however further data points are required to accurately determine K_i.

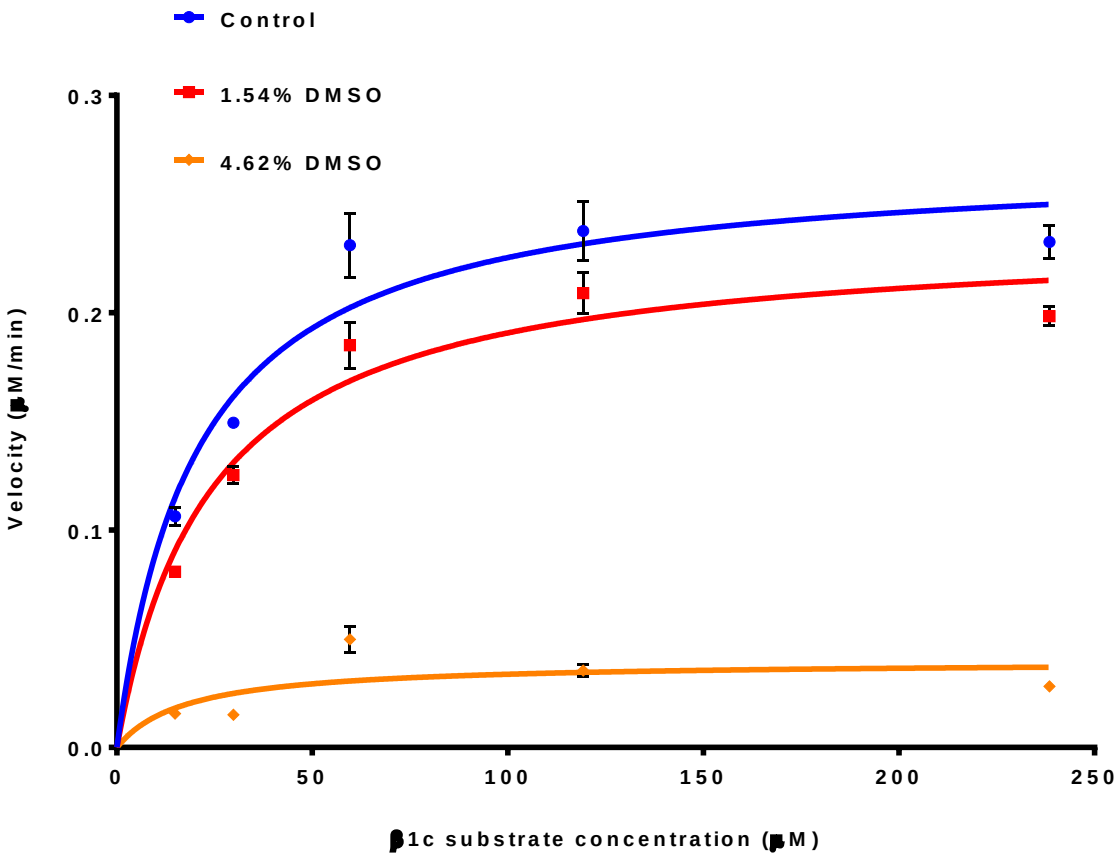
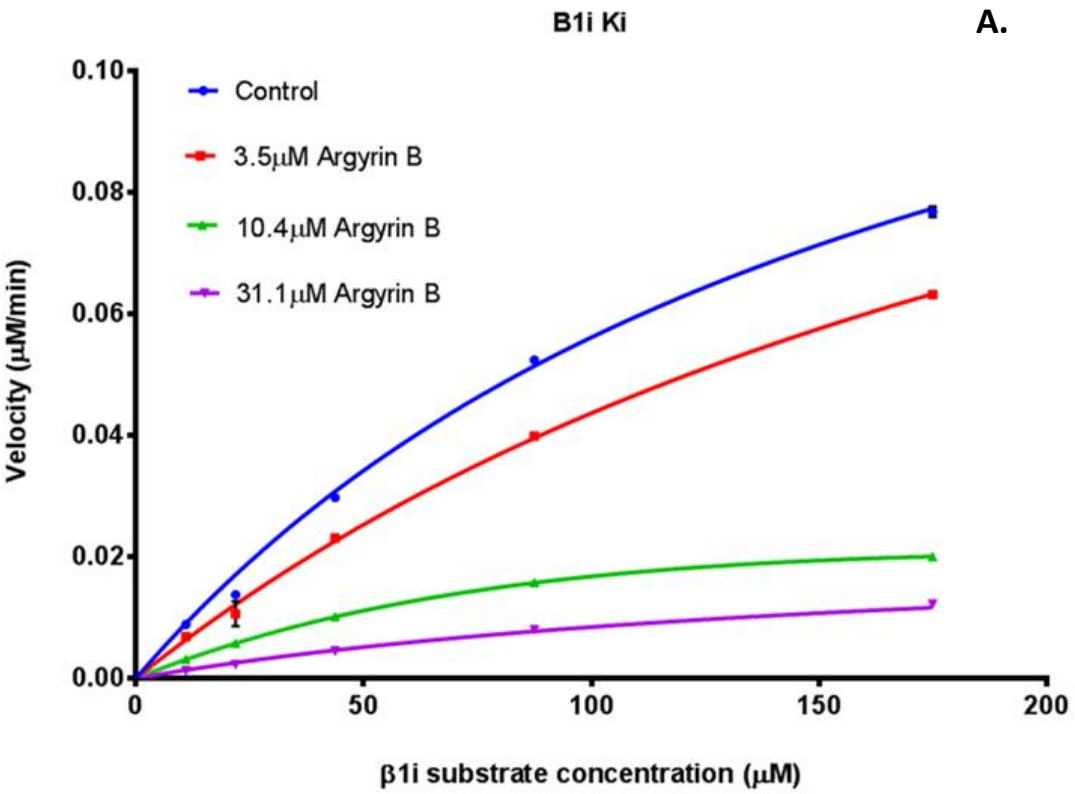
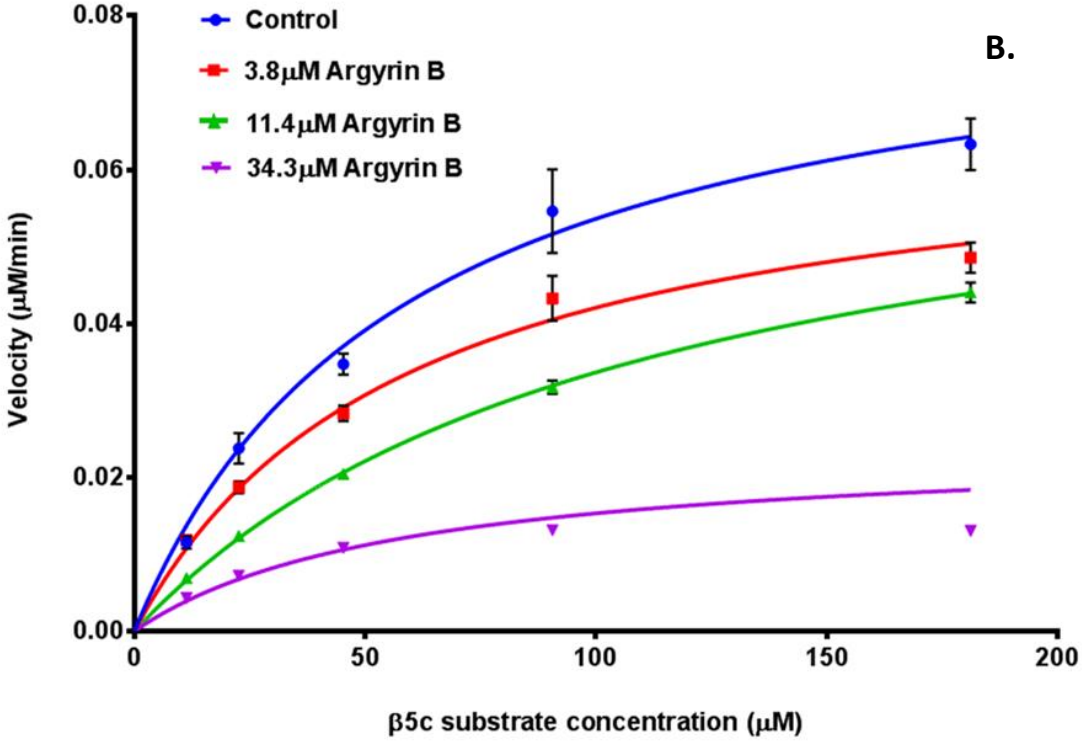


Figure 19 – β1c DMSO control data at high concentrations – DMSO concentration corresponding to argyrin B at 0 µM, 183.7 µM and 551.1 µM concentrations. Tested at [CP] = 0.1 µg/well, and range of Z-LLE-AMC concentrations. Based upon argyrin B preparations from from 10 mg/ml stock.

For data on β1i, β5c and β5i, best-fit model comparisons were first performed using Akaike's information criteria with results shown in supplementary table 7. The probabilities clearly disfavour competitive model fit against others, whilst comparison with un-competitive provided ambiguous and incomparable fits. In β1i and β5i, the non-competitive model was the preferred fit, whilst β5c showed marginal preference towards the mixed inhibition model. Further analysis using extra sum-of-squares F tests are shown in supplementary table 8. Contrary to the AICc test, the low p-value concludes that β5c showed statistically significant preference towards the non-competitive model. β1i and β5i best-fit outcomes associated with AICc tests and confirmed a suitable use of non-competitive analysis. As such to determine K_i values, non-competitive SNLR was used for all sets data shown in figure 20 and summarised in table 9.


B1i Ki
A.
Control
3.5μM Argyrin B
10.4μM Argyrin B
31.1μM Argyrin B
Velocity (μM/min)
0.10
0.08
0.06
0.04
0.02
0.00
0
50
100
150
200
β1i substrate concentration (μM)


B.
Control
3.8μM Argyrin B
11.4μM Argyrin B
34.3μM Argyrin B
Velocity (μM/min)
0.08
0.06
0.04
0.02
0.00
0
50
100
150
200
β5c substrate concentration (μM)

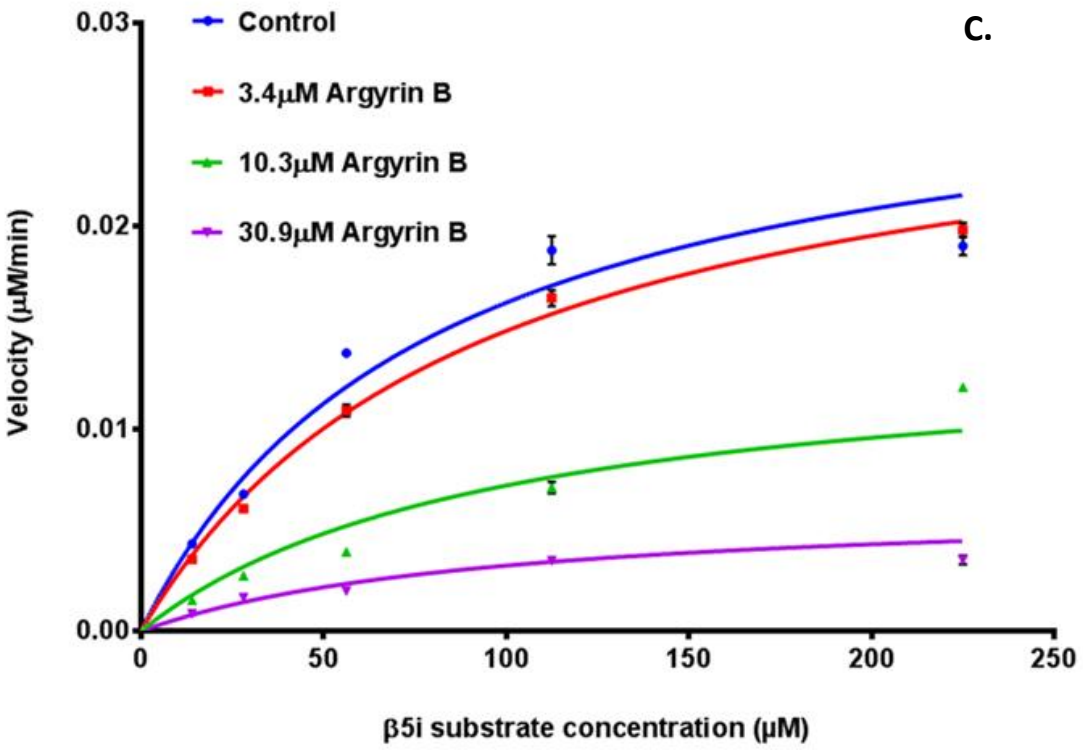


Figure 20 – K_i analysis at β1i, β5c and β5i active sites – All using 0.1 µg/well proteasome concentration and range of [argyrin B] at original IC50 x 0, 0.33, 1 and 3. [S] at K_m x 2.5, 1.25, 0.625, 0.1325, 0.15625 for each active site. Initial velocity determined and analysed by simultaneous non-linear regression using the best-fit inhibition model. Standard error bars from triplicates displayed for all data points. A) β1i, B) β5c, C) β5i.

Table 9 - Summary of K_i values at β1 and β5 sites of the CP and IP – Determined from non-competitive simultaneous non-linear regression analysis.

	β1c	β1i	β5c	β5i
Ki (µM)	**>100**	**5.2**	**13.9**	**6.6**
SE	-	0.34	0.96	0.49
95% CI	-	4.54 – 5.89	11.93 - 15.76	5.63 – 7.59
R-square	-	0.9639	0.9447	0.9127

K_i values reveal a greater effect of argyrin B at β1i and β5i, compared to other sites. However, statistical analysis could not be performed due to single replicate data available. This data correlates with IC_{50} estimations showing greater impact of argyrin B at the IP over CP. However, quantitatively K_i estimations differ from IC_{50} values previously determined (table 8) furthermore, β5i and β1i show different trends between IC_{50} and K_i estimations. Although K_i could not be specified at β1c, a high value was evident in comparison to all other active sites.

Further to SNLR analysis, alternative analysis methods were also tested with Cornish-Bowden and Dixon plots to estimate the inhibition type for β1i data. Supplementary figure 5 Dixon plot represents a mixed inhibition type that is close to non-competitive due to line intersections near the x-axis. Analysis of x-axis intersections between each line shown in supplementary table 4, revealed an average competitive K_i (K_{ic}) value of 4.08 µM +/- 0.31. Combined with a Cornish-Bowden plot shown in supplementary figure 6, a mixed inhibition is confirmed, again close to non-competitive. Supplementary table 5 shows each line intersection, generating an average un-competitive K_i (K_{iu}) value of 10.55 µM +/-0.73. These findings match those of the previous analysis, although final K_i values slightly differ.

In addition, a conventional Hanes-Woolf plot was used to determine K_m/V_{app} shown in supplementary figure 7 and subsequently plotted for each inhibitor concentration to estimate K_i in supplementary figure 8. This plot estimates $-K_i$ at the x intercept. Calculated from the line equation with y=0 as well as slope SE from least squares curve fitting analysis, a K_i of 4.14 µM +/- 0.26 was determined. Small error bars and similar values to SNLR analysis help to verify findings and analysis techniques used.

3.2 Computational modelling results

3.2.1 Sequence alignments

All substitutions from murine to humanised structural data were based upon Blastp alignments shown in supplementary figure 9-11 and sequence changes are summarised in supplementary table 9. Furthermore, sequences from structural data active site cuts were each aligned for similarity with percentage scores summarised in table 10.

Overall, relatively few amino acids are substituted from murine to humanised IP in each active site, at around 10%, with many of these changes conservative. There are no common substitutions observed between all active sites. At β2, many changes are towards the end of the chain, further from the Thr1 catalytic centre and its surrounding binding pockets.

Table 10 - Percentage sequence similarity between human, mouse and humanised CP & IP chains - pBLAST performed on sequence of active site cuts. β1in red, β2 in blue and β5 in green expressed as percentage values.

β1, β2, β5	Human CP	Human IP	Mouse CP	Mouse IP	Humanised IP
Human CP		62, 57, 70	97, 97, 97	64, 58, 72	62, 57, 68
Human IP	62, 57, 70		60, 59, 70	90, 89, 91	99, 97, 98
Mouse CP	97, 97, 97	60, 59, 70		63, 59, 69	60, 57, 69
Mouse IP	64, 58, 72	90, 89, 91	63, 59, 69		90, 89, 91
Humanised IP	62, 57, 68	99, 97, 98	60, 57, 69	90, 89, 91	

Protein sequence similarity searches reveal similar trends amongst conservation of β1, 2 and 5 across human, mouse and humanised IP & CP. Human and murine CP show near identical active sites, whilst human and murine IP show a little variance with approximately 90% similarity. Humanised IP sequences derived from structural data bear near perfect similarity to human IP sequences, with small discrepancies due to gaps.

To facilitate understanding of differences between the active sites modelled, complete structural alignment of human CP and humanised IP data are shown in supplementary figure 12. This includes type of substitution characterised by conservation of chemical properties. Amino acids of strongly similar properties score >0.5 in the Gonnet point accepted mutation (PAM) 250 matrix, whereas weakly similar properties are scoring <0.5. There are numerous amino acid differences between the CP and humanised IP at approximately 30-40%. This may result in overall chemical properties and structural characteristics of the active site. However, some residues pertain more to ligand binding affinity than others, dependent upon positioning around the catalytic centre and specificity pockets.

As well as sequence alignments, active site cuts were structurally aligned to observe any overall changes in positioning. Figure 21 part A represents an overview of the entire cut centred at Thr1 and including surrounding chains, whilst part B shows differences around the binding region of argyrin B.

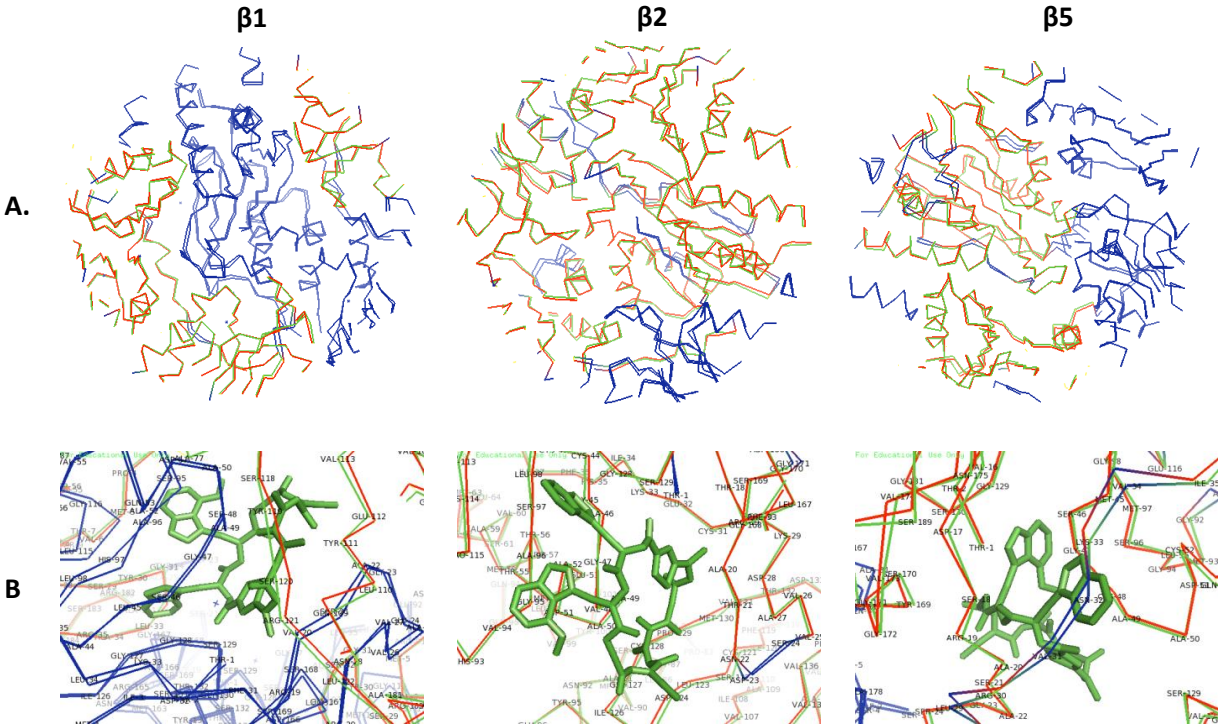


Figure 21 – CP overlaid humanised IP active site cuts – Human CP active site cuts in red, aligned to humanised IP active site cuts in green with differences highlighted in blue. A) Displaying entire cut, 28 Å from Thr1. B) Displaying area around the Thr1 catalytic centre, with overlaid argyrin B at proposed binding position from best-fit in humanised IP simulations.

From figure 21, β1 shows the greatest difference in structural positioning that is likely due to the P115 deletion in humanised β1i observed in supplementary figure 12 and conserved across various species (Huber *et al.*, 2012). In contrast, β2 and β5 remain relatively similar in structures of the active subunit chain. The difference observed in β5 overall cut is largely within a neighbouring subunit chain and is caused by a single deletion. Closer inspection around the active site reveals an overall shift between β1 chains, whilst β2 remains very similar and β5 has minor differences observed.

3.2.2 AutoDock molecular modelling

3.2.2.1 AutoDock summary

The binding affinity properties of argyrin B were recorded from the best-fit conformation of each 10 repeats, with results summarised in table 11 from raw data shown in supplementary table 10.

Results from 2-sample t-tests or non-parametric Mann-Whitney tests shown in supplementary table 12-14 and supplementary figures 13-15, reveal statistical significance between lowest binding energy for sets of data.

Table 11 – Predicted binding energies of argyrin B at each active site of the CP and humanised IP from AutoDock docking simulations – Argyrin B docked using Lamarckian genetic algorithm with 10 independent repeats, each of 50 runs. Statistical tests for normality and equal variance followed by mann-whitney or 2-sample t-test to determine significance. Red * = statistically significant to all other data sets. Green * is statistically significant to other green * only.

Binding properties	β1		β2		β5	
	Human CP	Humanised IP	Human CP	Humanised IP	Human CP	Humanised IP
Lowest Binding Energy (Kcal/mol)	-10.12	-11.83	-11.13	-11.05	-10.72	-10.97
Average Binding Energy (Kcal/mol)	-9.81*	-11.66*	-10.73	-10.69	-10.58*	-10.73*
Binding Energy SE of mean (Kcal/mol)	0.062	0.045	0.097	0.080	0.027	0.056
Average K_i (nM)	68.34	2.94	16.06	15.09	17.84	14.78

Argyrin B displayed a significant difference in binding energy preference towards β1i compared to any other CP & IP active site. Interestingly the CP counterpart β1c was the least favourable site for argyrin B binding, with a significant difference to all other sites. Argyrin B interactions at β2c and β2i were very similar in estimated binding energy, although showed the greatest variance in binding energies, reflected by higher SEM values. β5 and β2 interactions were similar, although a significant difference is observed between β5c and β5i, with β5i more energetically favourable. Average K_i estimations follow a similar trend to average binding energies.

The positioning and coordinates of each best scoring repeat were also analysed to measure reproducibility and consistency. Overlays of argyrin B are shown in figure 22, revealing the frequency at which the overall best score (highlighted in purple) adopted a similar position to all other replicate best scores.

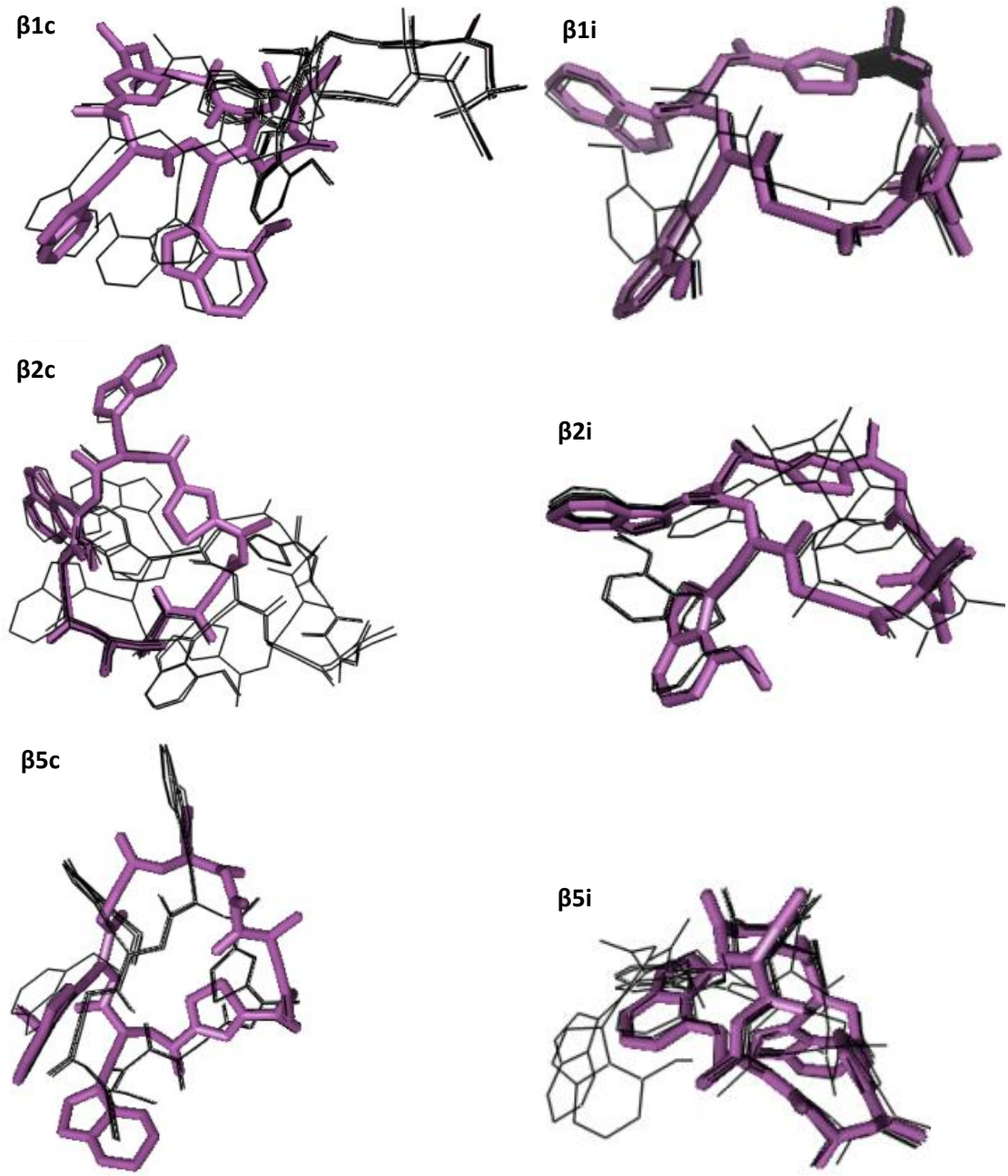


Figure 22 – Overlaid most energetically favourable argyrin B conformation of each 10 repeats – All tests from AutoDock docking simulations at human CP and humanised IP active site cuts. Overall best conformation presented in bold purple.

For β1c docking simulations, 2 major conformations were amongst the top scoring fits. These exhibited a 60:40 split and occupied significantly different areas of the active site. In contrast, β1i pose predictions showed 90% in a very similar position within the active site. Only β1 sites showed similar orientations of argyrin B between the CP and IP, although this is only true in terms

of shape and does not match in positioning or interacting residues. The β2 site followed a similar trend of the IP showing greater consistency of conformations; however orientations between β2c and β2i greatly differ. At the β5 site, argyrin B binds a similar area in all repeats, although the rotation of tryptophan groups and orientation differs both amongst repeats and between each proteasome. β5c displayed 2 main conformations almost equally, in which the argyrin is flipped with the tryptophans positioned on opposing sides.

Nonetheless, variations within the most favourable binding energy score were similar in repeats, evidenced by low standard errors, despite the different conformations and interactions. It should also be noted that only the top binding energy score for each is recorded, therefore the same conformation may well be identified, although not ranked as top.

3.2.2.2 Control TMC-95A

Pose predictions and lowest binding energy estimates from TMC-95A docking at yeast β2 sites, were compared to crystallography derived structural data. The overall upper quartile of binding energy conformations were recorded in supplementary table 15 . From these, 38% represent a close resemblance to crystal structure data, also producing consistent binding energies and interactions. These 38% were identified as the dominant conformational cluster from the 50 poses of each run, using an RMSD tolerance of 2 Å. The best predicted interaction bound to the Thr1, although the orientation differs to that suggested by crystallisation data, shown in figure 23. Whilst other favourable conformations deviated from specificity pockets and the catalytic centre with no particular trends or common positions identified.

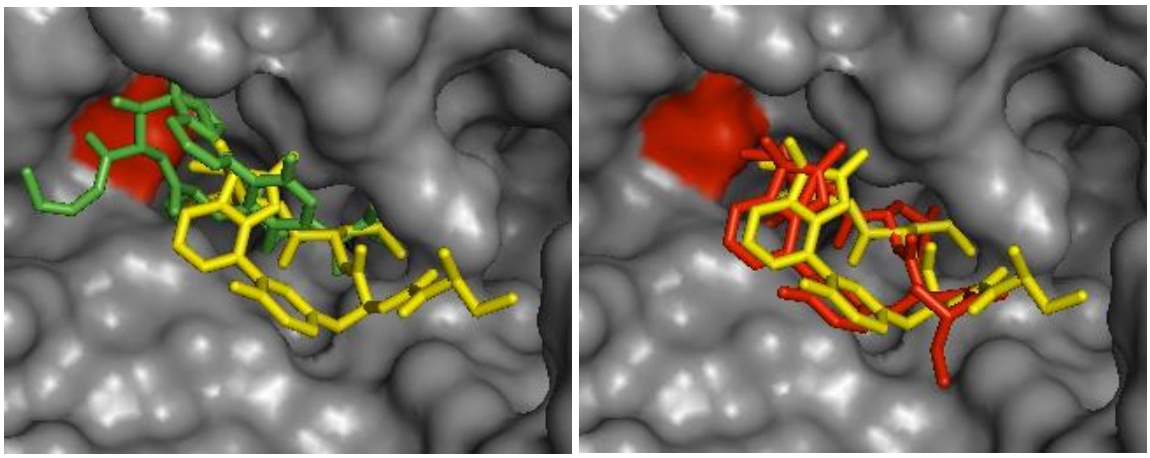


Figure 23 – Comparison of crystallisation data to AutoDock simulation predicted binding using TMC-95A at yeast β2c - Green represents the most energetically favourable pose against the crystallisation structural data in yellow. Red represents conformation similar to crystal data, that constituted 38% of the upper quartile ranked repeats. Surface colour red indicates position of Thr1 catalytic centre.

Overall, the most commonly predicted conformation shows similarity to structural data however, this is not always identified as the lowest binding energy. This highlights the importance of a high number of repeat tests and use of supporting laboratory data.

3.2.2.3 β1 site

From molecular docking simulations, the most energetically favourable conformation from each repeat was investigated in greater detail. The residues within the active site that interact closely with argyrin B were recorded, along with any hydrogen bonding. Findings of the frequency are noted in table 12 whilst the best scoring conformation is displayed in figure 24.

Table 12 - β1c and humanised β1i residue interactions from argyrin B AutoDock simulations – Best-fit of each 10 replicates analysed. Bold denotes presence in best conformational fit overall and italics denote residues in different subunit chain.

β1c Human		
Amino Acid Interactions	Occurrence /10	H bond occurrence /10
ARG35	10	**4 x single**
GLY47	10	6 x single
GLY97	10	
MET116	10	6 x single
MET95	10	
GLY129	7	**4 x single**
ALA49	6	
GLY23	6	6 x single
SER130	6	1 x single
SER46	6	**4 x single**
THR1	6	
THR20	6	
THR21	6	6 x single
THR22	6	
TYR30	6	
GLY128	4	
LEU33	4	
MET5	4	
PRO115	4	
SER133	4	
GLY31	2	
ASN3	1	
TYR134	1	

β1i Humanised		
Amino Acid Interactions	Occurrence /10	H bond occurrence /10
ALA49	10	
ARG19	10	
HIS97	10	
MET5	10	
SER129	10	**9 x single**
SER46	10	**10 x single**
SER48	10	
SER95	10	
THR1	10	**2 x single, 9 x triple**
VAL20	10	
GLY128	9	
GLY47	9	
SER21	9	**9 x single**
ALA96	6	
LEU115	1	
TYR30	1	

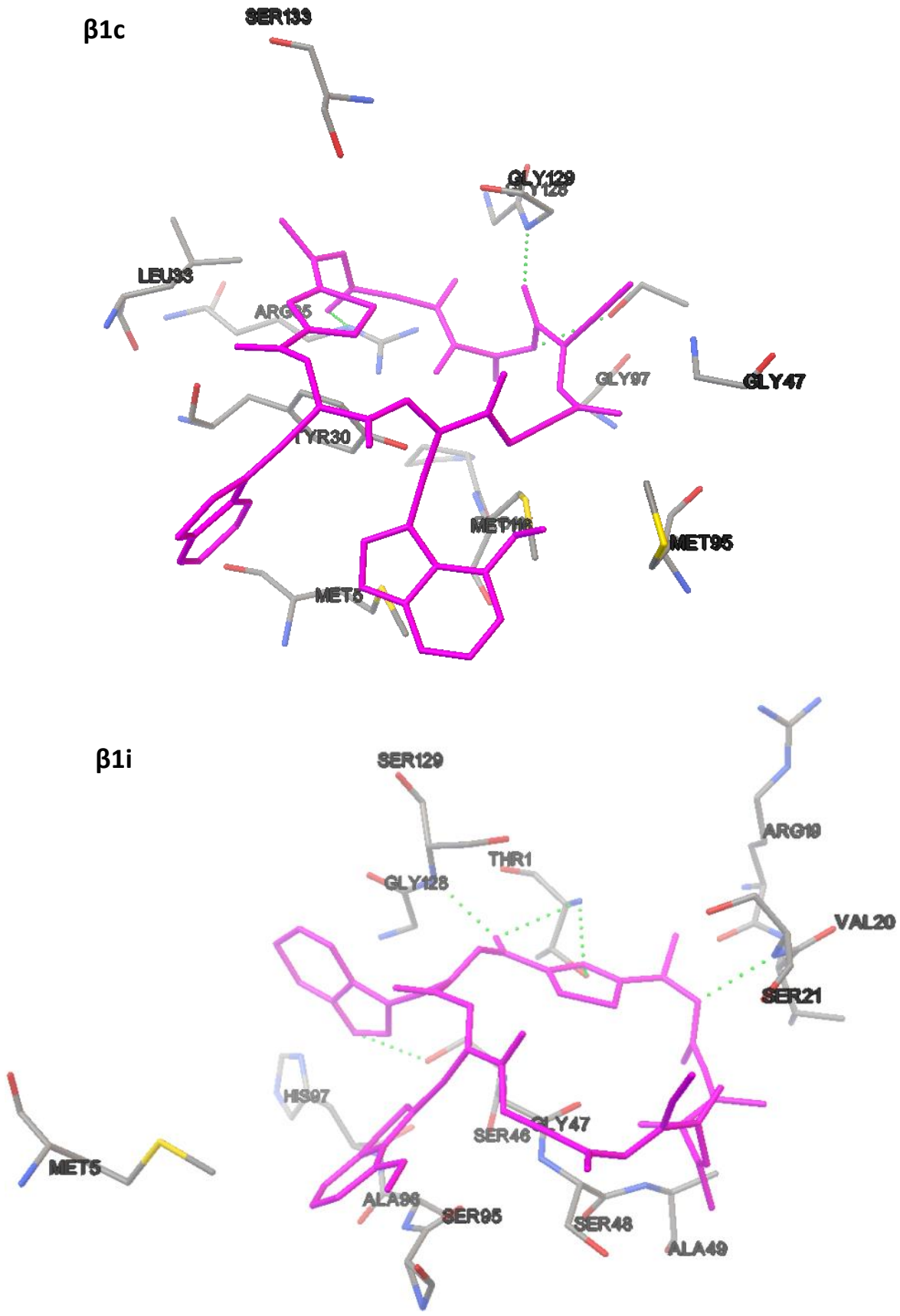
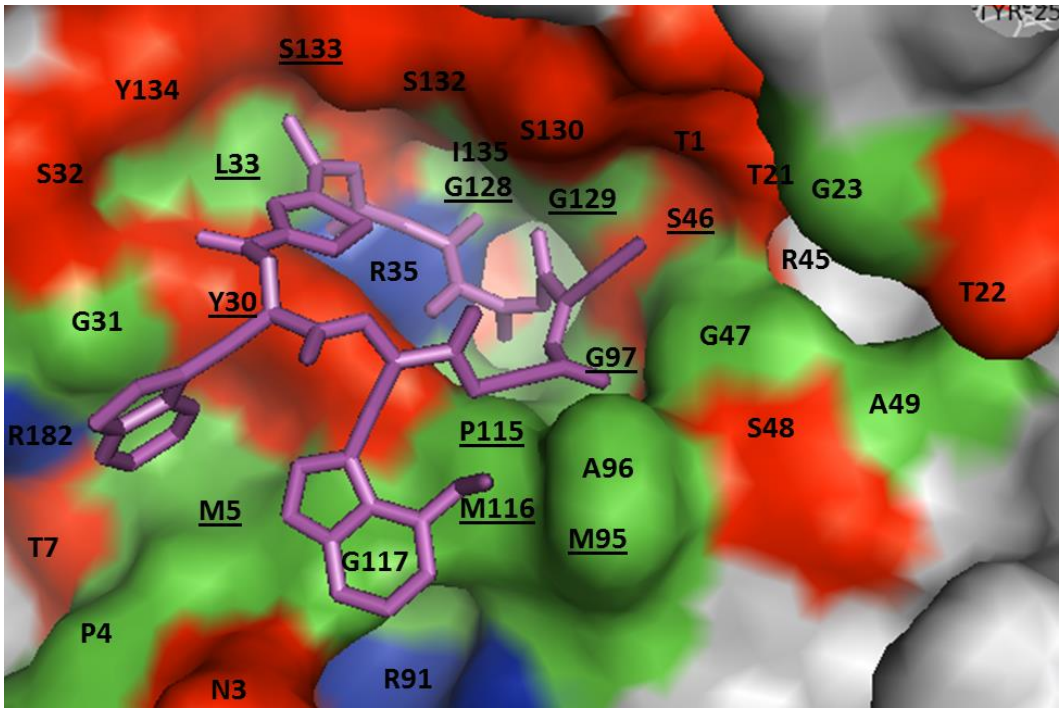


Figure 24 – 3D representation of all argyrin B interactions at β1c and humanised β1i sites from AutoDock predicted best conformation – Displayed residues are interacting with argyrin B whilst green dotted line represents H-bonding.

β1c



β1i

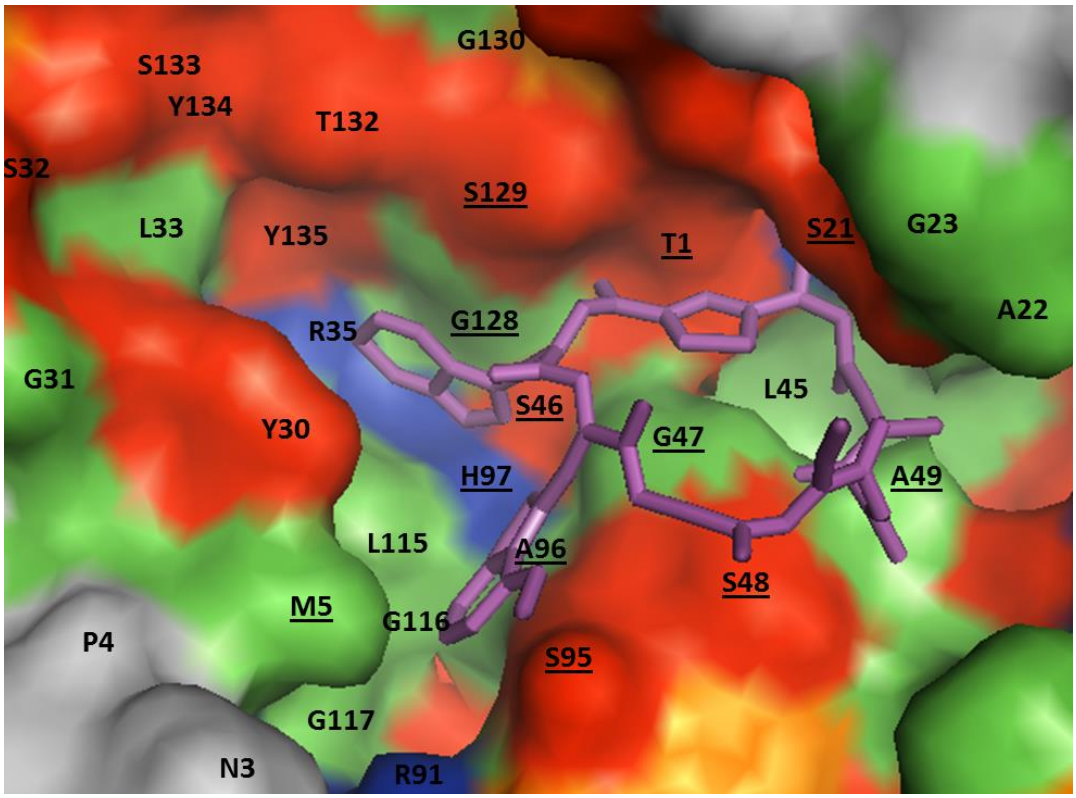


Figure 25 – AutoDock simulated argyrin B best-fit with surface representation at β1c and humanised β1i active sites – All nearby residues labelled and visible, interacting residues underlined. Coloured by general amino acid characteristics where red = polar, green = hydrophobic, blue = basic, orange = acidic.

Interactions at β1i showed greater consistency with 13 amino acid interactions and strong H-bonding around Thr1, observed in at least 90% of best-fit replicates. Likewise, many of the hydrogen bonds were between the same amino acid and part of argyrin B, on most repeats; also at a greater abundance compared to β1c. In contrast, β1c displayed 2 main conformations, therefore a greater variety of interacting residues and split frequency of hydrogen bonding. However, 5 core residue interactions were observed throughout, highlighting the importance of: Arg35, Gly47, Met116, and Met95 in argyrin B binding at β1c. Argyrin B also displayed interactions with residues from neighbouring and opposing chains from β1c. There are few interactions in common observed at both the CP and IP, affirming the possibility of selectivity.

In addition to the interacting residues, this information was also investigated with consideration of changes to the active site characteristics. A surface representation of best-fit conformation at each active site is shown in figure 25. In general, the β1i pocket appears to adopt more polar and basic characteristics, as well as different specificity pocket sizes compared to its constitutive counterpart. In β1i, argyrin B is positioned close to Thr1 and with part of the molecule interacting around the S1 specificity pocket. In contrast, at β1c a more unconventional position is adopted, away from the specificity pockets. Differences in key residues appear to significantly alter the shape of the active site, subsequently changing argyrin orientation.

3.2.2.4 β2c & β2i sites

β2 active site interactions from best-fit conformations of argyrin B molecular modelling repeats are summarised in table 13. Furthermore, figure 26 displays the orientation of binding interactions of the overall, most energetically favourable pose revealing sites of interaction on argyrin B.

Table 13 – β2c human and humanised β2i residue interactions from argyrin B AutoDock simulations – Best-fit of each 10 replicates analysed. Bold denotes presence in best conformational fit overall and italics denote residues in different subunit chain.

β2c Human		
Amino Acid Interactions	**Occurrence /10**	**H bond occurrence /10**
ALA97	10	
GLY47	10	1 x single
GLY95	10	
PRO115	10	
SER129	10	**7 x single,** 1 x double
MET127	8	
ALA96	7	
LEU132	7	
PHE33	7	
SER131	7	3 x single
THR126	7	
THR48	7	
TYR25	7	
TYR144	7	
GLY128	4	
THR1	3	3 x single
THR21	3	
ALA20	2	1 x single
ALA46	2	
ALA49	2	
ARG19	2	
GLY92	2	
TYR114	2	
ASP125	2	
ALA50	1	

β2i Humanised		
Amino Acid Interactions	**Occurrence /10**	**H bond occurrence /10**
ALA49	10	
ASP124	10	
GLY128	10	**7 x single**
THR1	10	2 x single, 6 x double, **1 x triple**
THR21	10	1 x single, 1 x double
ALA20	9	1 x single
ALA46	9	
GLY47	9	
ARG19	8	
GLN131	8	
GLY95	8	
PRO115	8	
SER97	8	2 x single
VAL48	8	
GLY92	3	
GLY45	2	2 x single
LYS33	2	
PHE33	2	
SER129	2	**8 x double**

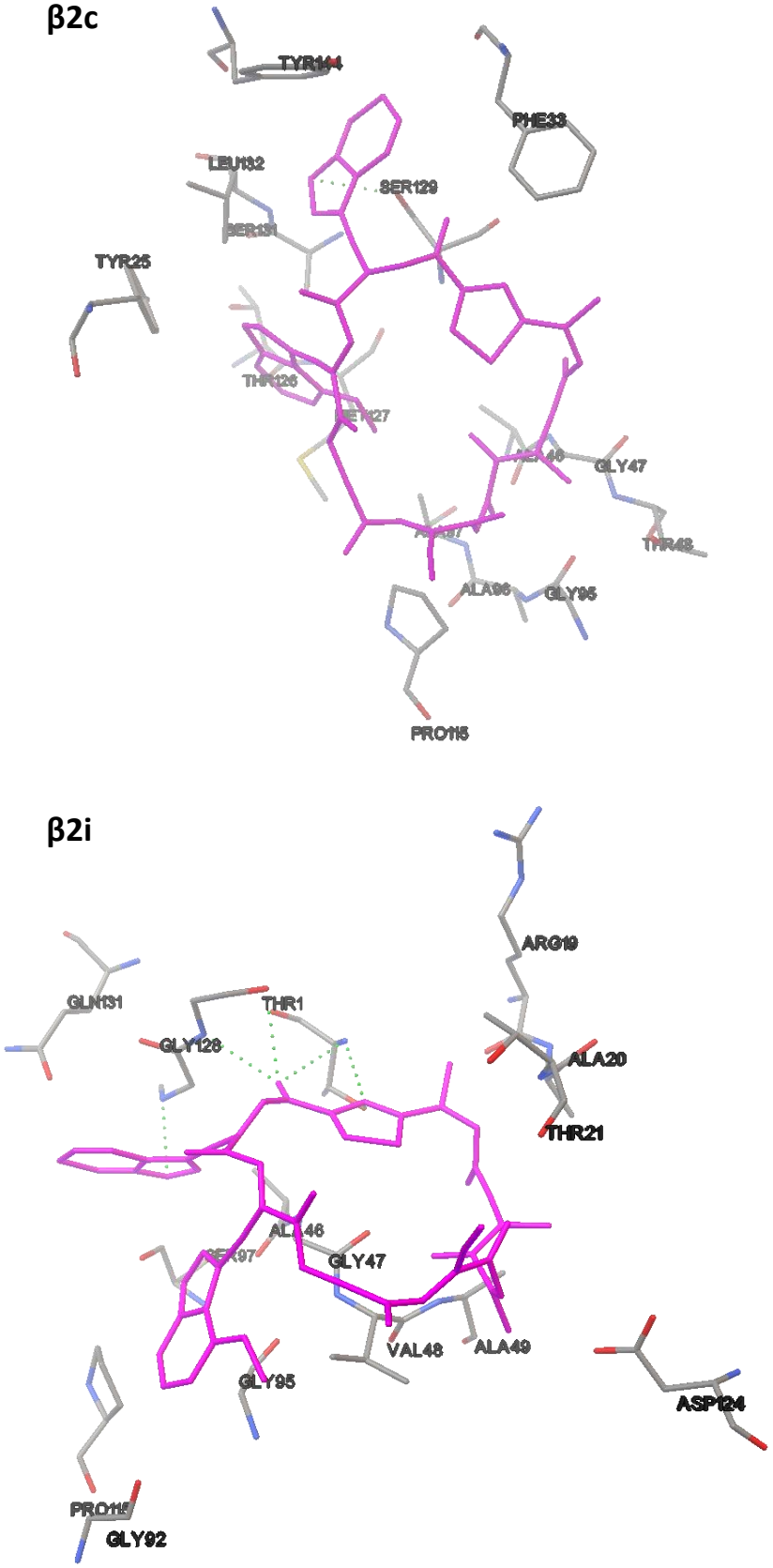
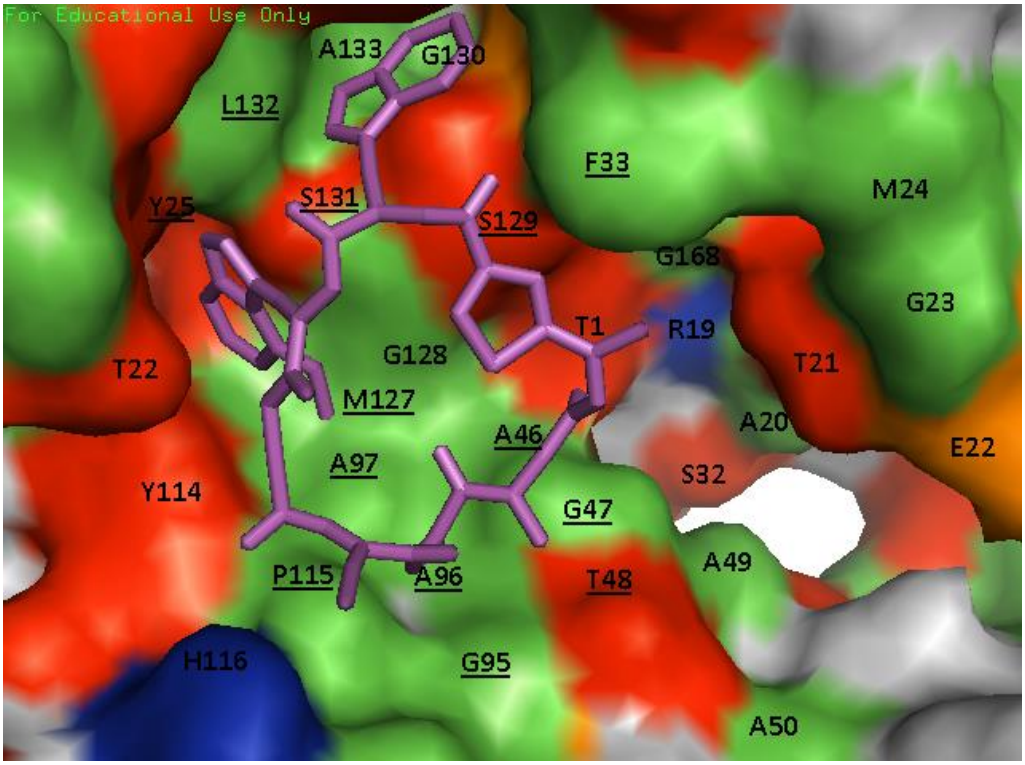


Figure 26 - 3D representation of all argyrin B interactions at β2c and humanised β2i sites from AutoDock predicted best conformation – Displayed residues are interacting with argyrin B whilst green dotted line represents H-bonding.

β2c



β2i

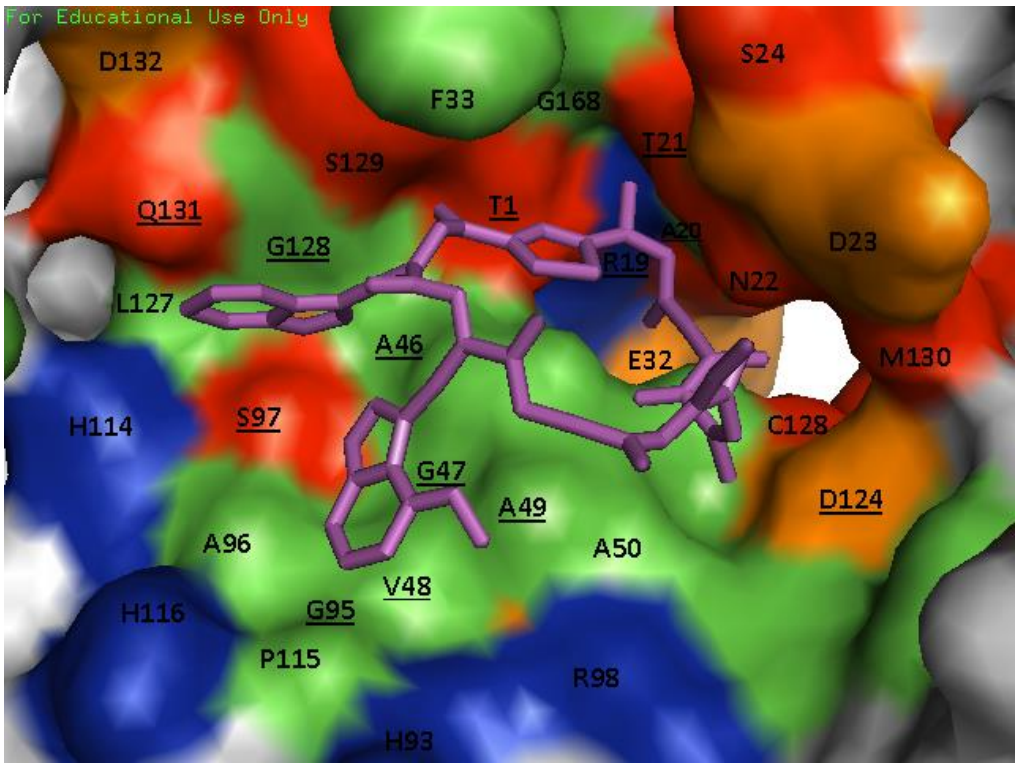


Figure 27 - AutoDock simulated argyrin B best-fit with surface representation at β2c and humanised β2i active sites – All nearby residues labelled and visible, interacting residues underlined. Coloured by general amino acid characteristics where red = polar, green = hydrophobic, blue = basic, orange = acidic.

β2c interactions are relatively consistent with at least 70% in common. Ser129 showcases important hydrogen bonding with the Trp moiety of argyrin B and some interactions are observed in neighbouring chains. Thr1 itself has minimal interaction at β2c, although the surrounding S1 pocket shows frequent intermolecular forces with the inhibitor. Within β2i a different orientation is observed with a high frequency of Thr1, Gly128 and Ser 129 hydrogen bonding. However, overall predicted binding energies suggest no significant difference in affinity between β2c and β2i.

Nevertheless, surface representations of best conformation within each active site, further reveal differences between site characteristics and architecture, shown in figure 27. The altered positioning of residues 22-24 (figure 21) affects the distance for argyrin B to interact and fit around the key Thr21 residue. The T48V and A97S non-conservative changes retain strong interactions in both fits; however the polar nature of S97 in β2i compared to the nonpolar A97 and the polar T48 in β2c compared to the nonpolar V48 in β2i, cause interactions at different areas of argyrin B, changing the orientation. Y93H changes from β2c to β2i appears to shift P115 and A96 that are each important in β2c binding. Furthermore, L132D and S131Q differences alter the shape and position of surrounding residues and reduces interactions at β2i 132 position.

3.2.2.5 β5c & β5i sites

Table 14 reveals the frequency of residue interactions from β5 site best-fit conformation repeats. All bonding interactions from the most favourable binding energy conformation are displayed in figure 28. In addition, figure 29 shows this within a surface representation of the active site.

Table 14 – β5c human and humanised β5i residue interactions from argyrin B AutoDock simulations – Best-fit of each 10 replicates analysed. Bold denotes presence in best conformational fit overall and italics denote residues in different subunit chain.

β5c Human			β5i Humanised		
Amino Acid Interactions	**Occurrence /10**	**H bond occurrence /10**	**Amino Acid Interactions**	**Occurrence /10**	**H bond occurrence /10**
ASN24	10	1 x single	**ALA49**	10	
GLU117	10		**GLY47**	10	1 x single
GLY129	10	**6 x single**	**MET45**	10	
GLY47	10	**8 x single**	**SER21**	10	**2 x single**, 7 x double, 1 x triple
SER130	10	**5 x single**	**SER46**	10	5 x single, **5 x double**
TYR113	10		**THR1**	**10**	**7 x single**
TYR169	10		**TYR169**	10	
THR1	9		**VAL31**	10	
ALA32	9		**GLY129**	8	
VAL128	8		**LYS33**	7	1 x single
SER96	7				
ASP115	6				
ASP167	6				
SER116	6				
TYR134	6				
SER23	5				
ALA46	4				
GLN33	4				
PHE137	4				
GLY98	3				
ILE30	2				
ALA22	1				
GLY99	1				

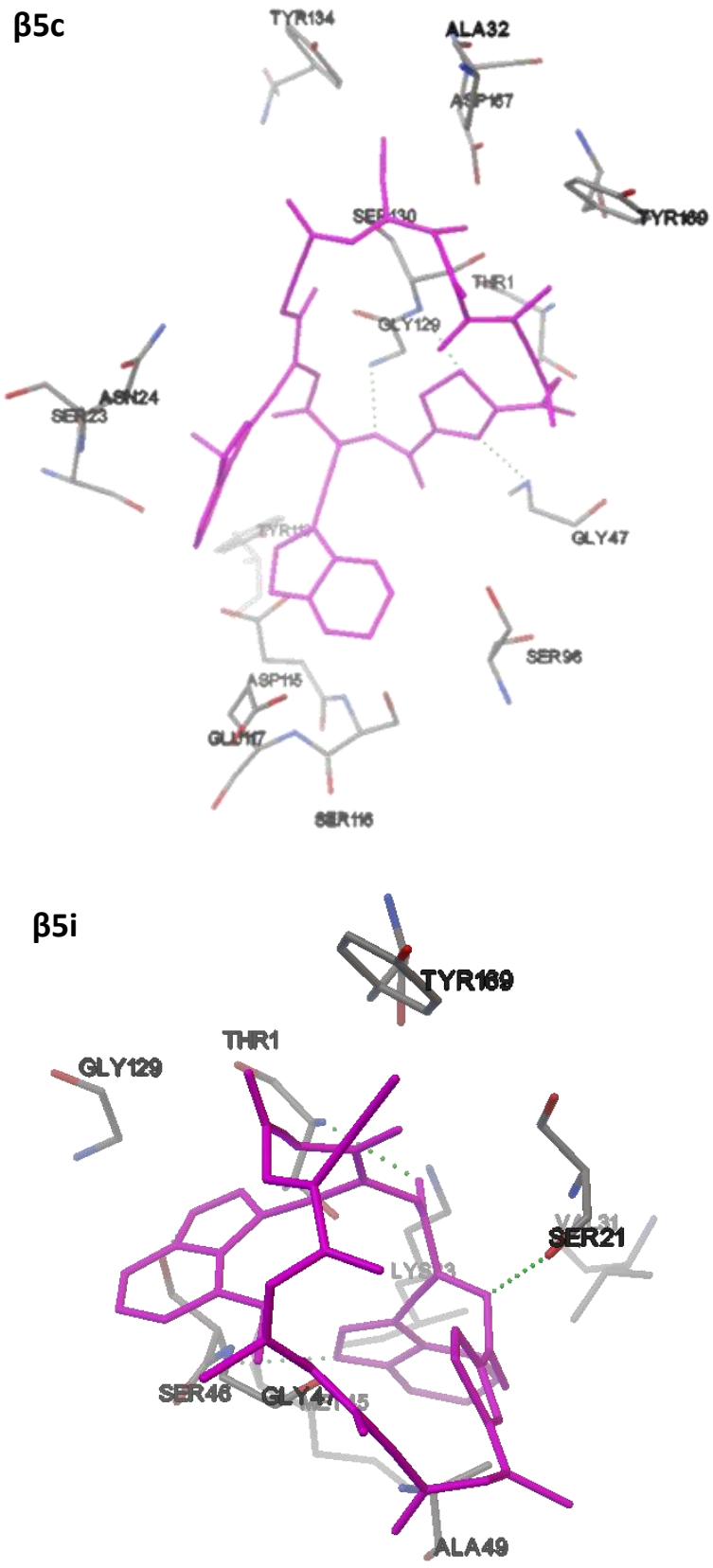
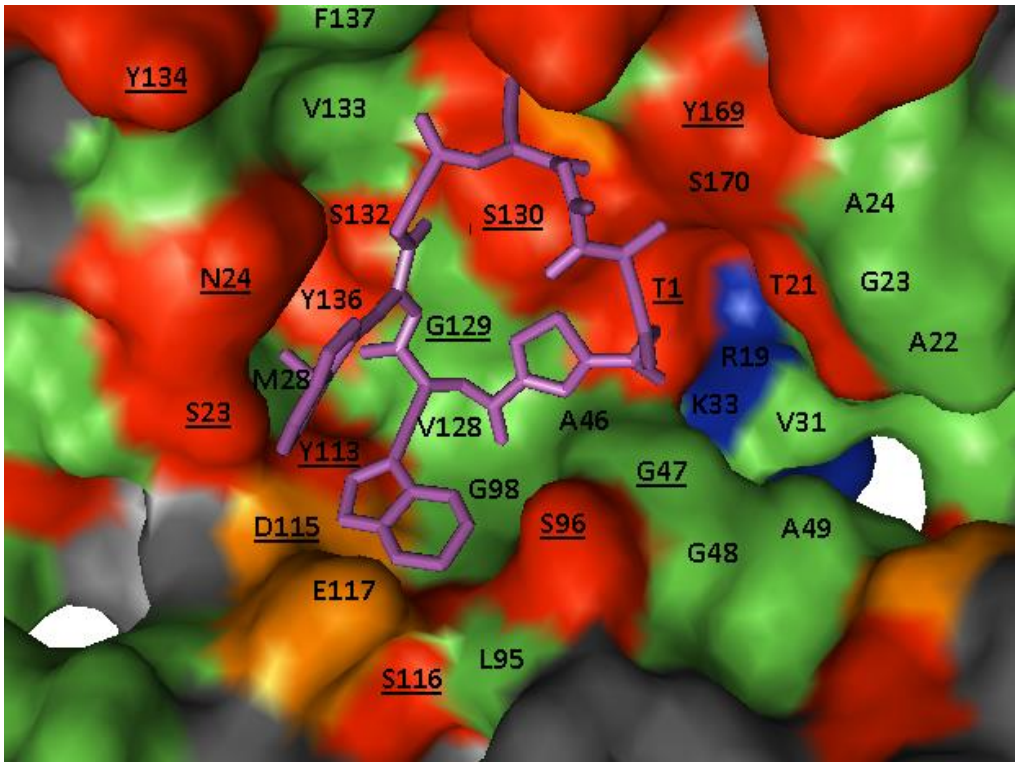


Figure 28 - 3D representation of all argyrin B interactions at β5c and humanised β5i sites from AutoDock predicted best conformation – Displayed residues are interacting with argyrin B whilst green dotted line represents H-bonding.

β5c



β5i

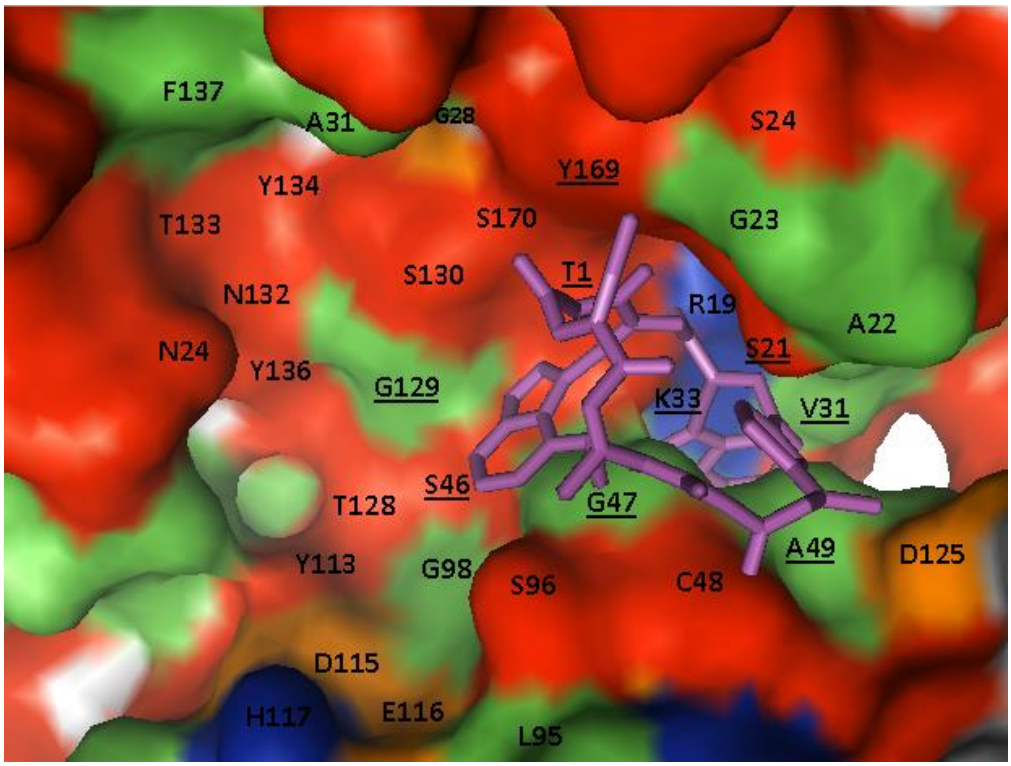


Figure 29 - AutoDock simulated argyrin B best-fit with surface representation at β5c and humanised β5i active sites – All nearby residues labelled and visible, interacting residues underlined. Coloured by general amino acid characteristics where red = polar, green = hydrophobic, blue = basic, orange = acidic.

There are 8 β5i site residues that interacted in all 10 best-fit conformations and only 10 unique interactions from all best-fit repeats. In contrast, β5c showed 15 residue interactions in the overall best-fit and 23 from the top 10. Despite more interactions at β5c, the hydrogen bonds and positioning in β5i predicted lower binding affinities. Despite displaying more consistent amino acid interactions, humanised β5i also showed a greater range of binding energy values. Through AutoDock binding energy estimations, β5i displayed a significantly favourable binding energy compared to β5c and conformations were distinct, highlighted by very few interacting residues in common.

The increased hydrophobic and polar nature of β5i (figure 29) is observed by key residue substitutions such as A46S, G48C and V128T. In B5i, all bonding is based around the Trp moieties whilst no residues strongly interacted at the opposing end of the inhibitor. The Trp rings of argyrin B wrapped around Ser46, each forming hydrogen bonds. Whereas, β5c showed hydrogen bond interactions from around the thiazole ring to Gly47 and Ser130. Argyrin B interacted with residues only from the β5 subunit in the IP, whereas conformations within the CP also displayed interactions with nearby subunit chains. Notably, Ser23 and Asn24 from the neighbouring β4 and Ala32 on the β3 from the opposing ring shown in the best conformational fit.

3.2.3 FRED Docking results

Pose predictions and interacting residues of argyrin B were tested using FRED, an alternative molecular modelling software. Ligand shape is treated as rigid, therefore at each active site, the resulting conformation and coordinates from the best-fit AutoDock simulations were tested. Resulting interactions are summarised in tables 15-17.

Table 15 – Comparison of argyrin B residue interactions at β1 and humanised β1i from AutoDock and FRED simulations – Comparison between single best-fit from all repeats, or frequency out of top 10 AutoDock repeats. Bold = hydrogen bonding, italics = different subunit chain and underscore = interactions unique to FRED docking.

β1c residue interactions		β1i residue interactions	
FRED with AD best-fit conformation	AutoDock best-fit	FRED with AD best-fit conformation	AutoDock best-fit
ARG35	***ARG35***	ALA49	ALA49
GLY128	GLY128	ALA50	
GLY129	**GLY129**	ALA96	ALA96
GLY31	2/10	ARG19	ARG19
GLY47	GLY47	GLY128	GLY128
GLY97	GLY97	GLY47	GLY47
LEU33	*LEU33*	HIS97	HIS97
MET116	MET116	*MET5*	MET5
MET5	*MET5*	**SER129**	**SER129**
MET95	MET95	SER131	
PRO4		SER168	
SER130	6/10	**SER21**	**SER21**
SER132		SER46	**SER46**
SER133	SER133	SER48	SER48
SER46	**SER46**	**SER95**	SER95
THR1	6/10	**THR1**	**THR1**
TYR30	*TYR30*	VAL20	VAL20
TYR134	1/10		
TYR136			
	PRO115		

β1c interactions closely match between FRED and AutoDock tests as well as hydrogen bonding with Gly129 and Ser46 remaining key sites. New interactions suggested from FRED docking of Pro4 and Ser132 neighbour other identified interactions at Met5 and Ser133. Likewise, Pro115 shown in AutoDock that is missing in FRED docking is situated adjacent to Met115, shown to interact in both. The same is observed in β1i, with the exception of Ser168 as a novel interaction predicted. Hydrogen bonding at Ser129, Ser21 and Thr1 are present in both simulations.

Table 16 - Comparison of argyrin B residue interactions at β2 and humanised β2i from AutoDock and FRED simulations – Comparison between single best-fit from all repeats, or frequency out of top 10 AutoDock repeats. Bold = hydrogen bonding, italics = different subunit chain and underscore = interactions unique to FRED docking.

β2c residue interactions		β2i residue interactions	
FRED with AD best-fit conformation	AutoDock best-fit	FRED with AD best-fit conformation	AutoDock best-fit
ALA27		ALA20	ALA20
ALA46	ALA46	ALA46	ALA46
ALA97	ALA97	ALA49	ALA49
GLY128	4/10	ALA50	
GLY23		ARG19	ARG19
GLY47	GLY47	ASN22	
GLY95	GLY95	*ASP124*	*ASP124*
MET127	MET127	ASP23	
PHE33	*PHE33*	GLN131	GLN131
SER129	**SER129**	GLY128	**GLY128**
THR21	3/10	GLY168	
THR22		GLY47	GLY47
THR48	THR48	GLY92	GLY92
TYR114	2/10	GLY95	GLY95
TYR144	*TYR144*	LYS33	2/10
TYR25	*TYR25*	PRO115	PRO115
	ALA96	**SER129**	**SER129**
	LEU132	SER97	SER97
	PRO115	**THR1**	**THR1**
	SER131	THR21	THR21
	THR126	**VAL48**	VAL48

β2c shows greater contrast with a number of interactions proposed by AutoDock, not established by FRED docking. The predicted FRED conformation appears to be a greater distance from Thr1, with argyrin B able to hydrogen bond the neighbouring chain Thr22. This may orientate the argyrin B out of reach to Ser131 and Leu132. Whilst Thr48 hydrogen bonding is predicted in both, Gly128 and Ser129 are suggested from FRED and AutoDock, respectively. β2i residue interactions show great similarities between each test. Although more interactions are suggested from FRED docking, these are neighbouring other residues predicted by both. Gly168 appears to be a distinct suggestion of an additional interaction and hydrogen bonding slightly differs.

Table 17 - Comparison of argyrin B residue interactions at β5 and humanised β5i from AutoDock and FRED simulations – Comparison between single best-fit from all repeats, or frequency out of top 10 AutoDock repeats. Bold = hydrogen bonding, italics = different subunit chain and underscore = interactions unique to FRED docking.

β5c residue interactions		β5i residue interactions	
FRED with AD best-fit conformation	AutoDock best-fit	FRED with AD best-fit conformation	AutoDock best-fit
ALA46	4/10	ALA20	
ALA32	*ALA32*	ALA22	
ASN24	*ASN24*	ALA49	ALA49
ASP115	ASP115	ARG19	
ASP167	ASP167	ASN32	
GLY129	**GLY129**	*ASP125*	
GLY47	**GLY47**	CYS52	
GLY98	3/10	GLY23	
MET28		GLY47	GLY47
PHE137	4/10	LYS33	LYS33
SER130	**SER130**	MET45	MET45
SER23	*SER23*	**SER21**	SER21
SER96	SER96	**SER46**	**SER46**
TYR113	TYR113	SER96	
TYR169	TYR169	SER24	
VAL128	8/10	**THR1**	**THR1**
VAL133		TYR169	TYR169
	GLU117	VAL31	VAL31
	SER116		GLY129
	THR1		
	TYR134		

AutoDock β5c interactions include Ser116 and Glu117 at the Trp moiety. However, FRED poses did not suggest such interactions, instead Met28 of a neighbouring chain and Val133 were predicted. Interestingly FRED docking did not include Thr1 as a major interaction, despite similar binding in nearby areas. β5i showed the greatest variation of all active sites in docking predictions. A number of additional interactions and an additional Ser21 hydrogen bond close to Thr1 are suggested from FRED docking. With the exception of the more distant Gly129, all AutoDock interactions and hydrogen bonds are matched in FRED predictions, yet FRED also predicts 9 additional interactions. Together, this suggests a potentially stronger affinity at β5i than AutoDock predicts.

Overall, many interactions and hydrogen bonding predictions between the docking software are shared. Those that are different between FRED and AutoDock are mostly neighbouring other shared interactions. Only β5i shows more of a variance and casts some doubt over the validity of the interactions listed by AutoDock.

Chapter 4. Discussion

4.1 Inhibitor kinetic studies

4.1.1 Determining enzyme activity

The use of peptide-AMC substrates has been an established method for determining proteolytic enzyme activity (Wildeboer *et al.*, 2009), including use in proteasome active site specific studies (Harris *et al.*, 2001; Bülow *et al.*, 2010). Kinetic assays to measure argyrin B inhibition were based upon measuring the rate of reaction. The properties of short peptides in active site specific substrates allow for binding and cleavage at only one active site, where β1c cleaves after the acidic glutamine of Suc-LLE-AMC, β1i hydrolyses Ac-PAL-AMC and β5 cleaves after hydrophobic tyrosine of Suc-LLVY-AMC (Miller *et al.*, 2013). However, it is important to note the range of substrates available (Kisselev and Goldberg, 2005). Whilst those used in this investigation are common and established, additional tests with alternative substrates would be beneficial to re-inforce active site specific activities.

The fluorescence of AMC was quantified by the electron excitation at 355nm, followed by a fall and release of light, measured at the emission wavelength 460nm. Coumarin fluorescence is quenched when bound to the P1 position amino acid of the substrate. However, upon cleavage of this peptide bond, the coumarin is released as shown in figure 30; allowing sensitive fluorescence increases to be continually measured.

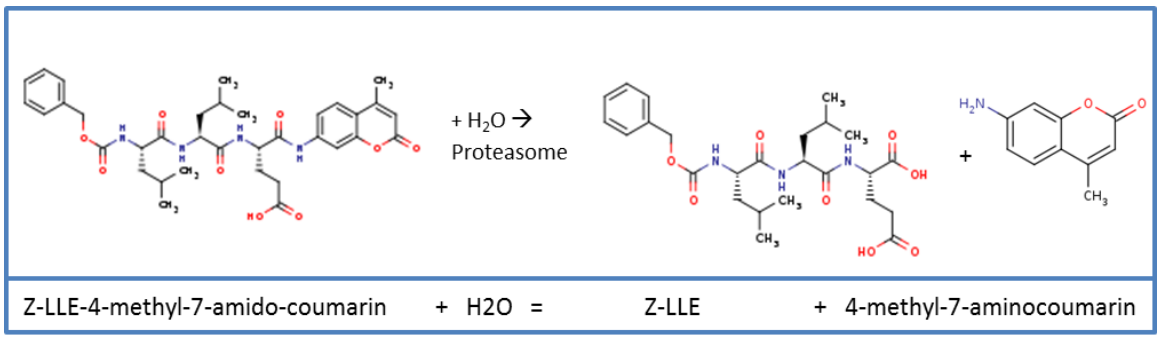


Figure 30 – β1c specific substrate reaction - Structure of β1 specific substrate Z-Leu-Leu-Glu-AMC, undergoing hydrolysis of peptide bond between carboxylic acid of P1 amino acid and the amino group of coumarin. The release of AMC is subsequently measured by fluorescence.

As a result, a substrate reaction with enzyme to form a single product is observed, as represented in equation 4. During initial stages of reaction the relative increase in [P] is greater than the decrease in [S] therefore; the accumulation of [P] is a more accurate measure. The molar concentration of substrate used was in excess of enzyme concentration, hence allowing sufficient substrate availability to bind free enzyme (E). The reverse reaction from product (P) to enzyme-substrate complex (ES) is not included as initial velocities were calculated whereby [P] is low and a steady state is assumed. Representative levels of E, ES and P are shown throughout a reaction in figure 31. This shows a pre-steady state as ES forms, followed by steady state of near constant ES at which point velocity is most appropriately measured (Copeland, 2005).

It is important to note that a combination of variables can contribute to the enzyme kinetics of a reaction therefore, strict parameters must be set. Any pH variation of H^+ or OH^- ions can alter charged groups of amino acids within the active site and cause conformational changes, hence affect reaction rate. Similarly, temperature changes cause instability of the enzyme, influencing rate (Copeland, 2005). Optimal conditions of pH 7.5 and 37 °C were used in accordance with physiological conditions, these also matched the vast majority of 20S CP investigations as summarised on the BRENDA database (www.brenda-enzymes.org, EC 3.4.25.1).

Equation 4 – Enzyme substrate reaction - E = enzyme, S = substrate, ES = enzyme-substrate complex, P = product, k = rate constant.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} P + E$$

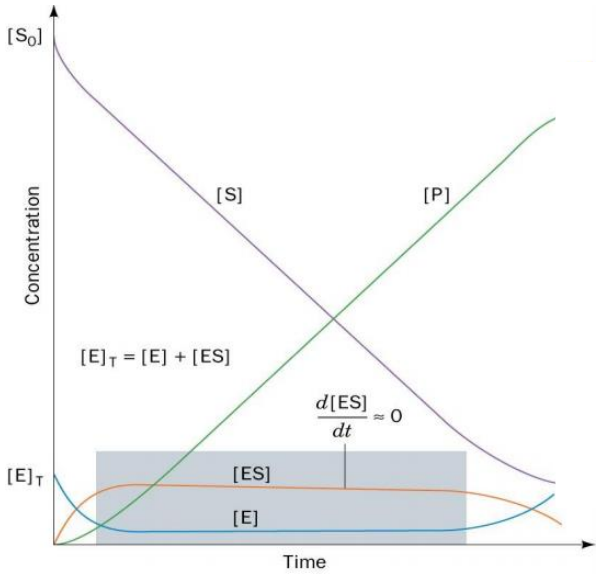


Figure 31 – Concentration of reaction components over time - Representative concentrations of free enzyme, substrate, enzyme-substrate complex and product concentrations in a reaction over time. Area of steady state assumption highlighted in grey.

Upon rate determination it is important to estimate K_m values before subsequent kinetic analysis, particularly as IC_{50} is measured at a single [S]. To determine IC_{50} values a [S] close to K_m is required to avoid unrepresentative values as observed in different inhibition modes, illustrated in figure 32; based upon theoretical values from Cheng-Prusoff equations (Cheng and Prusoff, 1973).

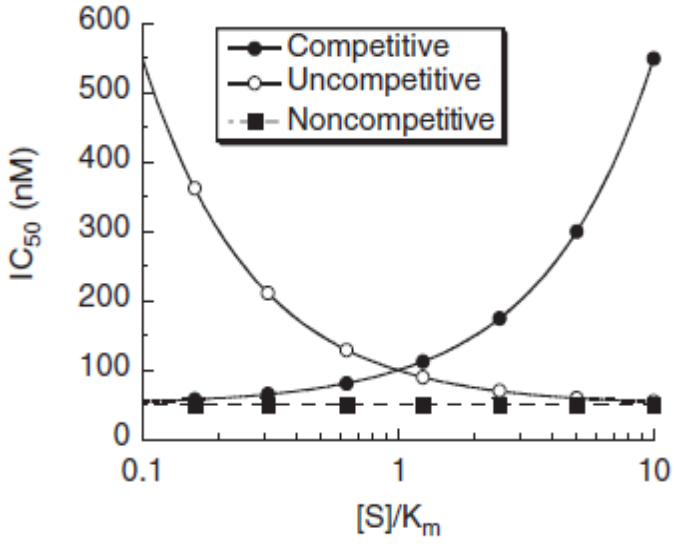


Figure 32 - Impact of substrate concentration on IC_{50} – Variation in IC_{50} values dependent upon deviation of substrate concentration from K_m values. Theoretical values based upon the Cheng-Prusoff equations in competitive, uncompetitive and non-competitive binding modes (Copeland, 2005).

The Michaelis-Menten equation and basis of non-linear regression analysis are derived from rate equations explained in supplementary figure 16. The series of equations facilitate understanding of how equation 4, representative of the reaction in this investigation, is quantified with kinetic constants using non-linear regression analysis.

It is the model in equation 10 (of supplementary figure 16) that is used in non-linear regression analysis to determine K_m, as well as forming the basis of models for different modes of inhibition and kinetic analysis. K_m represents the concentration of substrate that produces half of the maximal enzyme activity (Strelow *et al.*, 2012).

In this investigation, Michaelis-Menten graphs (figure 17) show apparent first order reaction with respect to substrate concentration, until higher concentrations when all enzymes are saturated with substrate. Therefore the rate is proportional to the concentration of substrate. Alternative methods using double reciprocal analysis of a Lineweaver-burk plot distort experimental error with lower concentrations producing a greater impact on the linear regression slope, evidenced in supplementary figure 3 analysis. Therefore non-linear analysis is generally regarded as more appropriate when a sufficient range of substrate concentrations are covered and this is indeed the prominent analysis method in enzyme studies (Kakkar *et al.*, 2000; Motulsky and Christopolous, 2003; Bisswanger, 2014). Nonetheless, some assumptions incorporated within regression analysis software are still questioned for accuracy and appropriateness (Cornish-Bowden, 2014).

Although figure 17 shows good fit and small SE bars, the 95% confidence level of K_m remain relatively broad. 95% confidence levels are created from SE multiplied t-distribution constant that is determined by degrees of freedom. The large range observed is therefore explained by lack of repeats whereby, when n=3, SEM is multiplied by 4.303, whereas with n=9, SEM is multiplied by 2.306. Overall, reasonable rates of reaction were produced under appropriate conditions and similar K_m values were obtained for each active site. This is expected since given an appropriate substrate, the proteasome is not known to have significantly greater efficiency at any particular active site (Kisselev and Goldberg, 2005).

4.1.2 Argyrin B IC_{50} values

Results from table 8 reveal lower IC_{50} values at IP active sites over their CP counterparts, in particular at β1i where a significant difference is observed. The IC_{50} represents the concentration

of inhibitor that causes 50% inhibition of the maximal activity (Mohan *et al.*, 2013) and is commonly used in pharmacological studies, including proteasome inhibition. Therefore, a lower value represents a greater functional strength of the inhibitor.

An appropriate range of concentrations are tested where values generally reach within 15% of maximal and minimal values, as well as spanning above and below the IC_{50} (Strelow *et al.*, 2012). An explanation for the curve plateau above 0% (observed in figure 18) may be due to insufficient mixing and time allowed for enzyme-inhibitor complexes to form, before the addition of substrate. Since sufficient data points were tested, a variable slope model was used as opposed to standard hill-slope of -1.0 of inhibitory dose-response curves. More negative hill slope values represent a steeper curve, observed in β5c, β5i and β1i of figure 18. An increased hill-slope value indicates a more dramatic change of inhibition over a narrow inhibitor concentration range. Whereas a low hill-slope value may be indicative of binding multiple active sites (Copeland, 2005), suggesting weak and off target effects of argyrin B at β1c. This is corroborated by a proposed non-competitive binding mechanism analysed in supplementary table 8. Existing, potent proteasome inhibitors typically display high hill-slope values around 1 due to covalent, competitive interactions (Lightcap *et al.*, 2000).

It is interesting to observe differences of argyrin B inhibition between each active sites of the CP, which has not been previously reported. Table 8 strongly suggest argyrin B displaying a greater functional binding strength at β5c over β1c with well characterised IC_{50} plots. This was tested using a wide range of argyrin B concentrations, substrate concentration equal to K_m and analysis of initial velocity data. In comparison, research by Bülow *et al.*, (2010) shown in table 18, displays percentage activity assays with 1.2 μM argyrin producing 52%, 55% and 58% remaining activity at β1, β2 and β5 sites respectively; suggesting similar argyrin B inhibition at each site. The approach from Bülow *et al.*, (2010) (table 18) that used a single inhibitor concentration, low substrate concentration and percentage activity measurement differs from this investigation methods outlined in section 3.1.2 & 3.1.3. The latter is considered a more appropriate kinetic measurement method (Copeland, 2000).

IC_{50} values are also reported by Bülow *et al.*, (2010), from MTT cytotoxicity assays in SW-480 colon cancer cells. This assay measures overall metabolic activity to reflect cell viability and therefore compound toxicity towards the cells. When compared to micromolar range IC_{50} values reported in table 8 from purified assays, an argyrin B IC_{50} of 4.6 nM (in table 18) suggests a substantial synergistic effect of each active site inhibition. Furthermore the long, 5-day incubation may

results in additional effects of the compound on other cellular processes, leading to increased toxicity and potency.

Table 18 – Argyrin analogue inhibition at constitutive proteasome active sites - Argyrin analogues A-F are tested in purified assays using 1.2 µM argyrin, 2 µg 20s CP and 50 µM site specific fluorogenic substrates at 37 °C and pH 7.8. IC_{50} values are determined from MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay measuring cytotoxicity following argyrin B incubation at 37 °C and 10% CO_2 for 5 days (Bülow *et al.*, 2010).

Compound	Remaining proteasome activity (%)			IC_{50} (nM) SW-480
	Caspase-like - β1	Trypsin-like - β2	Chymotrypsin-like - β5	
Argyrin A	29 +/- 5.2	45 +/- 4.5	30 +/- 6.0	3.8 +/- 0.3
Argyrin B	52 +/- 3.0	55 +/- 8.5	58 +/- 6.5	4.6 +/- 0.6
Argyrin C	38 +/- 8.3	40 +/- 7.0	43 +/- 5.5	1.5 +/- 1.1
Argyrin D	50 +/- 6.0	62 +/- 8.0	64 +/- 5.2	3.6 +/- 2.0
Argyrin E	65 +/- 4.0	70 +/- 3.0	70 +/- 5.5	520 +/- 270
Argyrin F	35 +/- 7.0	38 +/- 6.5	28 +/- 2.5	4.2 +/- 0.4
Argyrin G	60 +/- 6.5	75 +/- 3.0	65 +/- 4.5	63 +/- 55
Argyrin H	52 +/- 5.5	60 +/- 7.0	51 +/- 8.0	30 +/- 2

As shown in table 18, argyrin A is suggested to be a more potent inhibitor of the CP compared to argyrin B. Further studies have tested argyrin A inhibition at each active site, revealing similar potencies to bortezomib when measured as weight/volume (Nickeleit *et al.*, 2008). Tests were performed with site specific fluorogenic substrates at 20S CP and results are not directly quantified, although rough graphical estimates suggest argyrin A IC_{50} values of 121 µM, 38.3 mM and 68.18 µM at β1, β2 and β5, respectively. The lower inhibition at β2c corroborates with percentage remaining activity findings from (Bülow *et al.*, 2010), whilst concentration values are relatively high compared to findings for argyrin B in table 8, despite a suggested greater potency of argyrin A. Nevertheless, argyrin A displayed low nanomolar IC_{50} values at cell lines from MTT cell proliferation assays and potent antitumour activities (Nickeleit *et al.*, 2008). The structural differences between argyrin analogues are summarised in figure 12 that may account for differences in binding. Proteasome inhibitors typically display IC_{50} values at each active site in the nanomolar range, although non-covalent inhibitors are often less potent (Blackburn *et al.*, 2010; Duibella *et al.*, 2014; Sosič *et al.*, 2016). Indeed, control tests with epoxomycin and vinyl sulfone produced near complete inhibition at low, 1 µM concentrations.

At the IP, β1i and β5i reveal similar IC_{50} values. However, interestingly these are both lower than their CP counterparts, with statistical significance for β1i. At the time of writing, there is no literature available on argyrin A-F inhibition at the IP. A 16-fold increase in IC_{50} at β1c compared to β1i suggests strong selectivity and potential therapeutic application. Similar ratios and IP selectivity have been achieved with other inhibitor compounds, as later discussed in 4.3.2. However, the majority of studies are performed on β5 sites and comparatively few show selectivity towards β1. β5 has traditionally been considered the most influential active site, although recent studies suggest the importance is exaggerated (Kisselev *et al.*, 2006). Due to minor differences in structure and characteristics of β2 sites (Huber *et al.*, 2012), very few β2i selective compounds have been discovered or synthetically developed (de Bruin *et al.*, 2014), therefore β2 sites were not tested with argyrin B as priority, as potency is predicted to be similar in each.

Although IC_{50} is commonly reported alone in pharmacological studies, the addition of K_i data is considered valuable in determining mechanism and binding affinity (Copeland, 2005; Cornish-Bowden, 2014). The relationship between IC_{50} and K_i values is not always straightforward. The Cheng-Prusuff equation (Cheng and Prusoff, 1973) links these values for different inhibition modes as shown in equation 5, although use of direct conversion from IC_{50} to K_i is not widely appreciated as a reliable estimation (Copeland, 2005). As such, further kinetic analysis was performed to determine the mechanism of action and absolute K_i values.

Equation 5 - Cheng-prusoff equations – Equations displaying the link between IC_{50} and K_i values in competitive, non-competitive and uncompetitive inhibition modes, respectively (Cheng and Prusoff, 1973).

$$IC50 = Ki\left(1 + \frac{[S]}{Km}\right)$$

Competitive

$$IC50 = \frac{[S] + Km}{\frac{Km}{Ki} + \frac{[S]}{\alpha Ki}}$$

When $\alpha = 1$, $IC50 = Ki$

Non-competitive

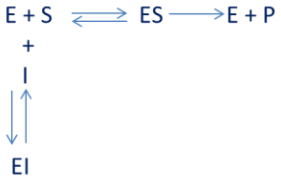
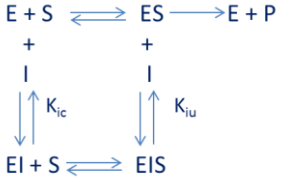
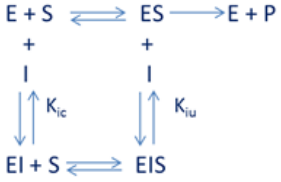
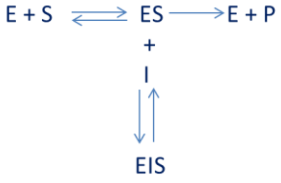
$$IC50 = \alpha Ki\left(1 + \frac{[S]}{Km}\right)$$

Uncompetitive

4.1.3 Modes of inhibition

The different modes of inhibition are summarised in table 19, showing where the inhibitor binds and its impact on K_m and V_{max} values.

Table 19 – Different modes of inhibition – Different types of inhibitor action, characterised by binding interactions and changes to Michaelis-Menten (K_m) and maximal velocity (V_{max}) values.

Type of Inhibitor	Description	Impact on K_m	Impact on V_{max}	Impact on K_m/V_{max} ratio	General Equation
Competitive	Inhibitor binds to free enzyme	↑K_m ↑[S] required to compete with I.	Same V_{max} Excess substrate can overcome I.	↑ K_m/V_{max}	
Non-competitive	Inhibitor binds equally to E and ES complex	Same K_m	↓V_{max} Always some E bound to I	↑ K_m/V_{max}	 $K_{ic} = K_{iu}$
Mixed	Inhibitor binds E and ES with different affinities	Variable K_m	↓V_{max} ESI complex can prevent P formation		 Mixed ratio of K_{ic} and K_{iu}
Un-competitive	Inhibitor binds ES complex only and cannot bind free E	↓K_m Some S bound to ESI therefore not converted to P.	↓V_{max} Always some E bound to I	Same K_m/V_{max}	

In competitive inhibition, the inhibitor binds only the free enzyme and a high concentration of substrate can overcome the inhibitor. This constitutes a high proportion of existing inhibitor drugs (Copeland, 2005). More rarely observed is uncompetitive inhibition where only the ES complex is bound by the inhibitor, reducing K_m and V_{max}. Alternatively, mixed inhibition occurs when the inhibitor is able to bind both the free enzyme and the ES complex, reversibly; quantified by Ki_c and

Ki_u values that can be linked by an alpha value. When α=1, (the inhibitor affinity for E matches that of ES), the inhibition is termed non-competitive. Alpha close to 0 suggests that the EI complex enhances substrate binding and the model is uncompetitive. Finally, a large alpha value indicates EI preventing substrate binding and a competitive mode of inhibition. Within literature, the use of non-competitive is often interchangeable with mixed inhibition, where exact alpha values of 1 are relatively rare (Mohan *et al.*, 2013).

In the presence and absence of inhibitor, comparing K_m and V_{max} values can help determine the mode of inhibition. In addition, graphical representations such as the Dixon and Cornish-bowden plots can be used (Cortés *et al.*, 2001). This is demonstrated in supplementary figure 5-6, each predicting mixed inhibition close to non-competitive, in accordance with SNLR results from supplementary table 7-8. It is not sufficient to simply use r-squared values to determine a best-fit nonlinear model (Spiess and Nuemeyer, 2010).

From Akaike's information criteria test, the model with the smaller AICc value is correct, although the difference between AICc values is reported in supplementary table 7; from that of the simpler model minus the more complicated model with increased parameters. As such, a positive difference in AICc shows preference for the more complicated model whilst a negative value represents preference for the simpler model (Kakkar *et al.*, 2000; Motulsky and Christopolous, 2003). Subsequently, probability values are reported based on the AICc difference. Competitive inhibition has fewer parameters than mixed that have fewer than non-competitive inhibition models. In each case a mixed or non-competitive model was preferred from argyrin B inhibition analysis.

In supplementary table 8 an extra sum-of-squares F test is used to distinguish between the nested models of mixed and non-competitive inhibition. This method tests with a p-value, therefore random scatter of data is taken into consideration. The F ratio defines the association between relative increases in sum-of-squares and degrees of freedom. An F ratio close to 1 suggests that the simpler model is more suitable, whereas greater than 1 favours the model with more parameters (Motulsky and Christopolous, 2003). For argyrin B at each active site, the non-competitive model was preferred, as a variant of mixed inhibition.

A non-competitive mode of inhibition was unexpected for argyrin B, with the majority of proteasome inhibitors characterised as competitive, including argyrin A (Nickeleit *et al.*, 2008) and F (Bülow *et al.*, 2010). The small molecule cyclic peptide is expected to block the active site as predicted in section 3.2 showing docking closely around Thr1, also shown in computational studies on argyrin A & F by (Stauch *et al.*, 2010; Loizidou and Zeinalipour-Yazdi, 2014). Although

allosteric interactions and ES binding are also likely, inhibition of the ES complex is difficult to determine with small peptide substrates. Cellular conditions of long peptide chain substrates may exhibit different affinities of argyrin B binding and steric blocking of binding sites surrounding the catalytic centre.

It is suggested that concentrations of 0.5 x K_m to 5 x K_m are used and higher concentrations of substrate facilitate the ability to distinguish between competitive and non-competitive inhibition. Results from figure 20 use substrate concentrations up to 2.5xKm due to limitations of plate reader reading limits and plate design. Some guidelines also suggest at least 8 inhibitor concentrations at each substrate concentration, ranging up to 10xK_i (Strelow *et al.*, 2012) whilst others suggest 4 up to 3 x IC_{50} (Copeland, 2005). The latter was followed in this experiment, however a greater range may prove useful. Despite this, many literature reports contain minimal range of concentrations tested, such as the 4 substrate and single inhibitor concentrations of argyrin F by (Bülow *et al.*, 2010), analysed with linear transformations.

As well as determining mechanism of action, K_i values were generated to support IC_{50} values.

4.1.4 Argyrin B K_i values

K_i indicates the potency of an inhibitor, measuring binding affinity through determining the concentration required for half maximal inhibition. Lower values represent tighter binding whilst higher values indicate weaker interactions (Mohan *et al.*, 2013).

Table 9 results for argyrin B at each active site reveal similar trends between the CP and IP, compared to table 8 IC_{50} data. In both cases, the K_i is lower at the IP site compared to CP counterpart, also observed to a greater extent at β1. The lower K_i is a valuable pharmacodynamic property associated with reduced toxicity. β1i shows the lowest K_i of all sites, in contrast to β5i displaying the lowest IC_{50}. However, in both cases neither were statistically significant from each other. Relatively high ranges of 95% confidence are observed for K_i estimations due to low number of degrees of freedom. Numerical values are relatively similar compared to IC_{50}, although these are expected to be near equal for non-competitive inhibition, as demonstrated in equation 5 (Cheng and Prusoff, 1973).

No experimental data of argyrin B K_i values at the CP or IP were available at the time of writing. However, (Bülow *et al.*, 2010) estimates the K_i of argyrin F at 81nM, 112nM and 76nM for β1c, β2c and β5c active sites, respectively. These are noticeably lower than reported argyrin B values

and do not show the same weak binding at β1c. Although it is important to note that argyrin F Ki values were determined from only a single 120 nM inhibitor concentration with 3 low substrate concentrations of 40 µM, 20 µM and 10 µM. Due to the similar structures and characteristics, kinetic properties of argyrin B were expected to be similar to its analogues, however minor structural changes in key areas have been shown to have a significant simulated effect on binding (de Bruin *et al.*, 2014; Loizidou and Zeinalipour-Yazdi, 2014).

It is important to note that whilst an inhibitor may express great affinity, tight binding, to the enzyme, this does not necessarily translate to efficacy, the biological response. Cell based assays and *in vivo* studies would provide valuable data building upon kinetic parameters.

Overall, interesting trends have been observed with argyrin B inhibition at each active site. Notably a preference towards IP active sites, in particular β1i. This selectivity shows great therapeutic potential for treatments associated with lower toxicity and fewer side effects as discussed in section 1.4.3. Findings from computational molecular modelling can postulate reasons for the increased affinity through analysis of predicted binding conformations, interactions and energetics.

4.2 Computational modelling and structure-based drug design

4.2.1 Structural data

Recent findings contribute towards a growing number of residues deemed important for inhibitor binding. In figure 33

Figure 33, sequence alignments ascertained from structural data shown in supplementary figure 12 are presented with additional information regarding specificity pockets, displayed with colour coding (Stauch *et al.*, 2010, Huber *et al.*, 2012, de Bruin *et al.*, 2014). This information facilitates the identification of key amino acid differences between the CP and IP sites that may result in the greater IP affinity exhibited by argyrin B. Furthermore, novel structure based drug design is no longer always focussed on Thr1 targeting, with success of alternative non-catalytic targets such as Cys48 of β5 (Duibella *et al.*, 2015).

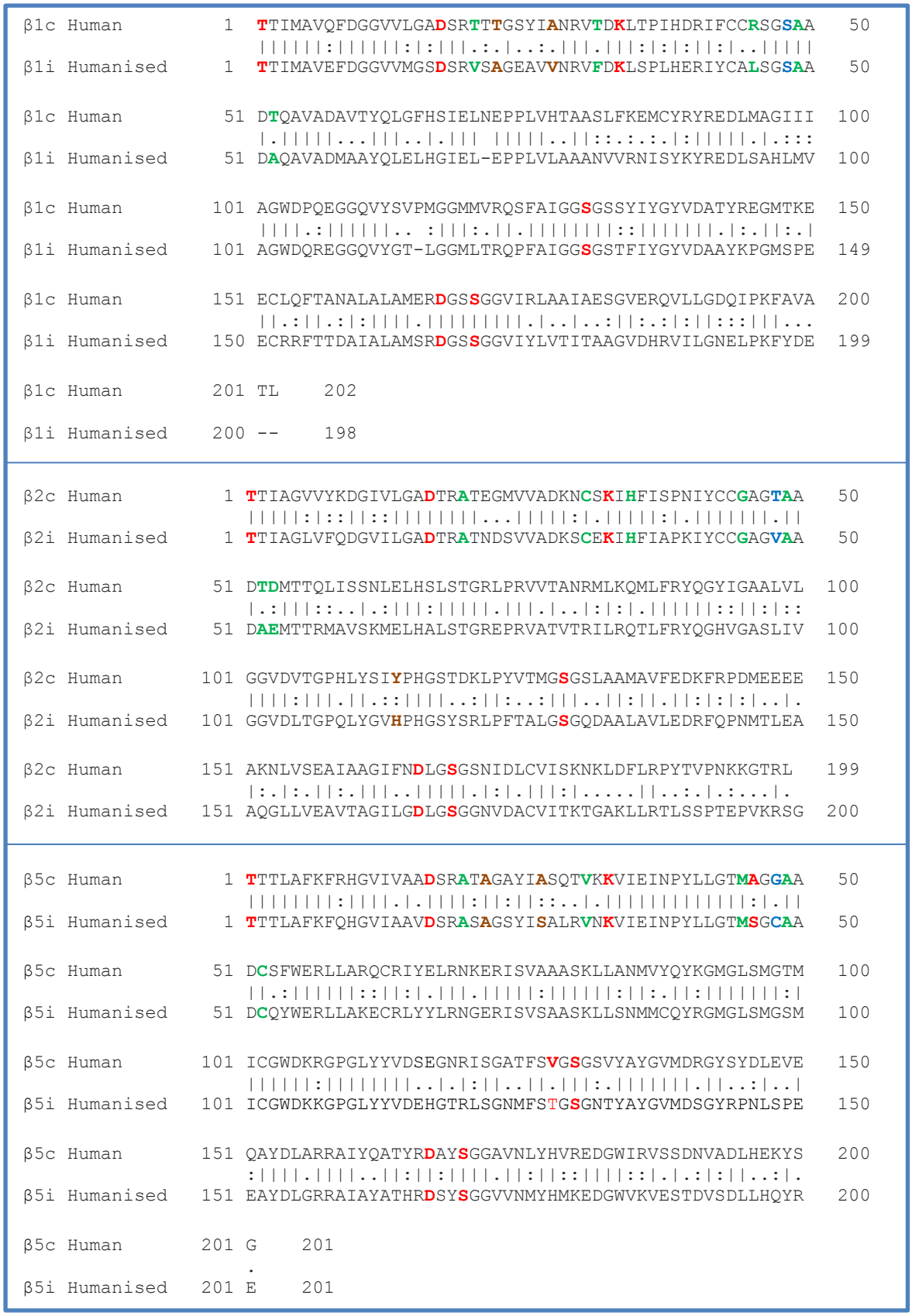


Figure 33 - Sequence alignment of human CP and humanised IP active sites with key binding residues highlighted - '|' = full conservation of the same residue, ':' = strongly similar properties and '.' = weakly

similar properties. Residues contributing towards substrate specificity pockets are coloured as: green = S1, blue = S2, brown = S3, red = importance for active site (Adapted from: Huber *et al.*, 2012).

Comparisons between active sites are first discussed then later covered in greater detail with respect to argyrin B binding in section 4.2.2.

Comparing the β1c and humanised β1i sites, numerous differences can be observed at the S1 pocket, namely: T20V, T31F, R45L and T52A, none of which retain strongly similar properties (de Bruin *et al.*, 2014). The S3 pocket also shows variation, with contrasting T22A and A27V from CP to humanised IP (Huber *et al.*, 2012). Several other residues of importance to the active site remain conserved, although β1i humanised lacks a residue at 115 for which β1c has a proline present which may cause a shift in the chain resulting in structural differences. None of the substitutions made in preparation of humanised β1i from murine β1i (supplementary table 9) involved amino acids known as important for the active site or specificity pockets. Although alternative substitutions may alter architectural arrangement, this suggests high sequence conservation amongst the main specificity pocket amino acids, as previously described in (Huber *et al.*, 2012).

Between β2c and humanised β2i, there are few residue differences within active sites and specificity pockets. Only T48V of S2 and T52A as well as conservative D53E of S1 are noted as different. T52A was a difference also observed between murine β2i and human β2i sequences, revealing potentially key species differences of the IP. Beyond the catalytic centre, there are many differences in the overall sequences that can affect the chemical property and structural characteristics of each site.

β5c and humanised β5i appear remarkably conserved amongst key residues. However, A46S and V128T from β5c to β5i humanised, differ which are suggested to modulate the active site. The previous M31V substitution from murine β5i to humanised β5i further highlights the species differences (supplementary figure 11).

In addition to focus on the active site subunit chain, the surrounding beta subunits have also been shown to influence inhibitor binding (figure 26). By producing active site cuts, there were overlaps into neighbouring chains, which in some cases proved influential to the overall architecture as well as interactions observed.

CP and IP active site sequence alignments of multiple organisms has also been tested by (Huber *et al.*, 2012), showing high levels of conservation between human, murine, zebrafish and rat species, amongst several more. Interestingly, many key changes between CP and IP counterpart sites are

also conserved, suggesting that different species models retain strong relevance in translation to human application.

The highlighted residues in figure 33 are noted in literature as being relevant in ligand and inhibitor binding, with potential for structure-based drug design to enhance selectivity (Blackburn *et al.*, 2010; Miller *et al.*, 2013; Huber *et al.*, 2016). Although, in the case of non-covalent, inhibition by the cyclic peptide argyrin B, additional interactions and conformations predicted in table 12-14 may also play key roles.

4.2.2 Argyrin B binding at constitutive and immuno- proteasome

4.2.2.1 β1c & β1i sites

Purified assay IC_{50} results shown in figure 18 reveal a significantly greater strength of argyrin B binding at β1i compared to β1c. This is further corroborated in K_i estimations figure 19-20 and molecular docking binding energy predictions table 11. Analysis of the positioning and interactions predicted can postulate structural reasons for the greater binding affinity observed.

The β1c caspase-like site typically cleaves after acidic residues, whilst β1i shows a more trypsin-like activity, cleaving after basic residues (Pellom and Shanker, 2012; Johnston-Carey *et al.*, 2015). This is due to different size and characteristic of the binding pockets, which appear to also impact the affinity of argyrin B. Argyrin B at β1c showed more variation with interactions, compared to β1i where a particular fit was evident. At β1c, the 2 different conformations shown in repeats suggest no clear pocket for strong interactions.

R45L substitution in β1i creates a more hydrophobic region at the base of the S1 pocket; where the positively charged arginine in β1c that would not attract argyrin's hydrophobic region, is replaced. This is observed with carfilzomib (structure in figure 11) where the hydrophobic P1 leucyl group is disfavoured at the polar β1c S1 pocket (Harshbarger *et al.*, 2015). Further substitutions T20V, T31F, T52A, replace the polar threonines in β1c with hydrophobic residues that are positioned at the S1 pocket. Of these, argyrin B only interacts directly with Val20, although there are increased interactions around the S1 pocket area with neighbouring residues. To potentially strengthen these interactions, hydrophobic and branched residues could be extended around the base of the cyclic peptide inhibitor. As also suggested in literature stating bulky, hydrophobic P1 residues as prerequisites for β1i selectivity (de Bruin *et al.*, 2014).

Furthermore, the T21S conservative substitution in β1i allows consistent hydrogen bonding at this residue (figure 24), not observed in β1c perhaps due to the smaller size of serine and proximity to S1 (de Bruin *et al.*, 2014). Loizidou and Zeinalipour-Yazdi (2014) also report lack of T21 bonding of argyrin A (structure figure 12) at humanised β1c, although, Stauch *et al.*, (2010) suggest T21 hydrogen bonds across all humanised active sites in argyrins A-D & F.

The T22A and A27V substitutions in β1i create a smaller S3 pocket (Huber *et al.*, 2012); however argyrin B shows little interaction at this site in either favourable conformation. This is contrary to reports by Stauch *et al.*, (2010) identifying dispersive interactions at these residues for argyrins A-D & F in humanised β1c. G97H in β1i creates a more basic and protruding residue which differentially accommodated the orientation and interaction of argyrin B's bulky tryptophan rings. Key interactions with Trp1 are shown whilst also orientating the N of Trp1 to allow H-bonding with nearby Ser46 (figure 25). Furthermore, M95S creates a polar region for Trp2 O to form dipole-dipole interactions. In contrast, β1c shows few interactions at either Trp moiety which may partly explain an overall lower binding affinity.

Lack of Val114 in β1i may have affected the chain positioning compared to β1c, as illustrated in figure 15. This subsequent shift of spatial arrangement may favour the proximity and ability of argyrin's thiazole ring to hydrogen bond multiple times with Thr1, only observed in β1i. This spatial arrangement may also explain the consistent hydrogen bonds observed in β1i at the polar Ser129, whereas β1c contains a neutral glycine at this position.

Whilst selectivity towards the β1i over β1c active site is sought after, the additional interactions at β2i and β5i are also important. Kisselev *et al.*, (2003) shows how caspase-like active site selective inhibitors can stimulate trypsin-like activity and allosterically inhibit chymotrypsin-like activity, therefore understanding affinity and interactions across all active sites is paramount.

4.2.2.2 β2c & β2i sites

Despite neither site displaying an energetically favourable bind with argyrin B (table 11), the different conformations and interactions (figure 26) suggest the possibility of developing specificity. This is interesting since the substrate binding pockets of β2c and β2i have been reported as difficult to distinguish and shallow in size (Huber *et al.*, 2012), with only 1 β2i selective inhibitor reported to date (Koroleva *et al.*, 2015).

Some substitutions within the active site result in distinct conformations of argyrin B between

each, leading to the possibility of synthetically developing selectivity. Key interactions are observed around residues 46-49, 96-97 and 128/129 in both β2 sites. Several of these residues have suggested importance of binding in literature (figure 33)(Huber *et al.*, 2012). In addition, argyrin B interacts at residues 19-21 and Thr1 of β2i at a greater frequency than β2c, perhaps due to E22N and G23D differences from β2c to β2i. Although not directly involved as interactions, these appear to shift the positioning of Thr21. Furthermore G23D and M24S in β2i may create steric hindrance and inhibitor stabilisation discrepancies between β2c and β2i. These larger and more polar residues in β2i are in close proximity to the S1 pocket and appear influential. Indeed, Stauch *et al.*, (2010) report E22 in humanised β2c to play a role in coordinating the polar indole H of argyrin Trp2 ring into a hydrophobic side pocket.

T48V and A97S from B2c to B2i, within key regions of interactions, may affect argyrin B positioning. S97 in β2i creates a polar region amongst largely hydrophobic residues. This subsequently favours interactions with Trp1, in some cases forming hydrogen bonds. Residue 48 plays a role in the S2 binding pocket (Stauch *et al.*, 2010; Huber *et al.*, 2012) and although Thr and Val have similar composition, an additional hydroxyl on Thr results in polar properties. Both are key for argyrin B interactions, although at different areas of the compound.

T52A and D53E from β2c to β2i are positioned around the S1 pocket, accounting for substrate recognition (Huber *et al.*, 2012; Koroleva *et al.*, 2015). However, these do not appear to interact with argyrin B or play any significant role in affinity. Similarly, Y114H in β2i, of importance to the S3 pocket, does not directly impact argyrin B binding. However this substitution may influence positioning of P115 in each active site, subsequently increasing interactions with argyrin B at β2c, whilst being out of reach at β2i.

Recently, Koroleva *et al.*, (2015) have reported exquisite β2i over β2c selectivity achieved by bisbenzimidazole derivatives measured by computed binding energy values and kinetic data. As observed with argyrin B, monomeric bisbenzimidazole (structure in figure 34) did not interact with β2c S1 pocket 53 residue either. However, the dimeric form, through an aromatic stacking folding conformation was able to interact at the β2i S1 pocket. These large compounds were often predicted to dock across large portions of the β2 and β3 sites and on occasions adopt allosteric binding (Koroleva *et al.*, 2015).

Overall, the differences of interactions and conformations balance similar final binding energies for argyrin B at each β2 site (table 11). Upon identifying different predicted poses and interactions, as well as novel developments of β2 selective compounds, it would be interesting to perform kinetic assays of argyrin B at β2c and β2i.

4.2.2.3 β5c & β5i sites

Docking simulations showed statistical significance of argyrin B binding at β5i to a greater affinity than β5c (table 11). Different interactions and conformations were predicted that can be attributed to specific changes in the active sites. It is important to note that whilst values from kinetic studies suggested the same trend of IP preference in IC_{50} (figure 18) and K_i (figure 20) values, these were not shown to be statistically significant.

Argyrin B adopted a more complete fit within the β5c subunit, displaying interactions with residues around most regions of the cyclic peptide and hydrogen bonds often forming from the thiazole ring (figure 28). In contrast, within β5i fewer residue interactions were observed with most occurring around Trp1 and Trp2, including hydrogen bonds. This suggests that binding could be further enhanced by modifications to argyrin B that extend at R_1 and R_2 of argyrin B in figure 12, to allow closer proximity to the subunit.

Argyrin B shows multiple interactions with β5i chain at positions 45-49, whereas β5c only interacts at Gly47. Although, Gly47 is conserved in the CP and IP, residues 46 and 48 are substituted. The CP contains hydrophobic Ala46 and Gly48 which are each smaller in size and less polar, compared to hydroxylic Ser46 and sulfhydryl containing Cys48 in the IP. An increase in specifically hydrophobic interactions has been previously shown to enhance inhibition at β5c for the natural phenol, epigallocatechin gallate analogues (Kanwar *et al.*, 2010); identifying this modification as influential to enhance binding. Together with A27S, the G48C substitution in β5i results in a smaller and more hydrophilic S2 pocket (Huber *et al.*, 2012). This can contribute to possible dipole-dipole interactions at Trp2 in the β5i conformation that are not observed in β5c. Aforementioned substitutions may also explain the alternative positioning of argyrin B whereby the thiazole hydrogen bonds to Gly47 of the CP. This is similar to Loizidou and Zeinalipour-Yazdi (2014) showing argyrin A hydrogen bonding between Gly47 of yeast β5c and the macrocyclic backbone. In contrast, within the best conformational fit of β5i, the Ser46 forms a double hydrogen bond bridging each Trp moiety. The tryptophans rotate to align δ^+ hydrogens bonded to the electronegative nitrogen of the indole ring, which each bond to the Ser46 (figure 28). Whereas, the hydrophobic Ala46 did not display interactions at β5c. Furthermore, Ala49 of the S1 pocket, which contributes to chymotrypsin-like activity, is conserved and shows consistent interactions in β5i but none in the corresponding CP subunit. Interestingly, Loizidou and

Zeinalipour-Yazdi (2014) identified hydrogen bonds of argyrin A at yeast Ala49 of β5c but not the humanised counterpart.

At β5i the configuration of Met45 increases S1 pocket size, assisted by the S53Q substitution in β5i. The larger side chain of glutamine is able to interact with Met45 in β5i, however the serine in β5c cannot attract Met45 to the same extent (Huber *et al.*, 2012). The larger β5i S1 pocket is then able to accommodate a larger P1 side chain as reported with other inhibitors (Parlati *et al.*, 2009; Harshbarger *et al.*, 2015; Sosič *et al.*, 2016). In the β5i conformation of argyrin B the aromatic Trp1 interacts in this position, enhancing binding energies.

The T30R and K32N substitutions from β5c to β5i have very different properties. The large, basic amino acids essentially switch positions either side of the Val31, between each subunit. This affects the overall architecture between β5c and β5i, thereby allowing the basic Lys33 and hydrophobic Val31 of β5i to interact around S1. This appears to be a key interaction in β5i with argyrin B, whilst β5c does not exhibit any interaction with this part of the active site.

Thr1 is an important interaction at both subunits and often displays hydrogen bonds in β5i with the O of the amide bond between Trp1 and Trp2. This strengthens the multiple interactions around the tryptophans of argyrin B within β5i. Whereas, at β5c, the positioning of argyrin B is such that only weak interactions are formed with Thr1 which are at the opposing end of argyrin B compared to β5i. This unconventional conformation instead resulted in Trp1 interactions at the negatively charged region 113-117 and Trp2 with alternative chains in the CP, notably Ser23 and Asn24 from β4c. These additional intermolecular interactions may help contribute to the overall binding energy in β5c, partially explaining the small overall difference compared to β5i fit where argyrin interacts with fewer residues overall.

Ser21 is another important interaction in β5i, often creating 1 or more hydrogen bonds between the thiazole and Trp2 of argyrin B. The 21 residue position is important, potentially influencing positioning of the conserved Ala20 and Ala22 either side that are key residues for the S1 and S3 pockets, respectively. The β5i T21S substitutions are of similar polar properties, but serine is slightly smaller without the extra methyl group, potentially allowing extra space for argyrin B at β5i. However, Huber *et al.*, (2012) also suggests S3 of β5i accommodates smaller, polar residues compared to bulky, hydrophobic side chains at S3 of β5c. Furthermore, Stauch *et al.*, (2010) and Loizidou and Zeinaipour-Yazdi (2014) identify T21 to occupy bulky Trp moieties of argyrins in humanised and yeast β5c. Beck *et al.*, (2015) evidence the importance of S3, through crystallographic analysis of non-covalent inhibition as a result of solely S3 occupation of β5c, again utilising T21 in hydrogen bonding.

The aromatic, polar Tyr169 interacts in all conformations for both subunits, although in different locations on the inhibitor. Positioned before the key Ser170 this is an important interaction (Huber *et al.*, 2012). Gly129 is also located adjacent to a key residue, Ser130. Gly129 interacts in both subunits however; in β5c this can form hydrogen bonds together with Ser130 towards argyrin B. This appears to be an influential interaction in β5c, appearing in only the favourable binding energy conformations. Ser130 is not observed to interact with β5i and is positioned around the Trp rings. The V128T from β5c to β5i may also influence Ser130 binding, with the valine a common interaction in β5c although interestingly not in the best observed conformation.

Overall, the residues of importance and substitutions from β5c to β5i shown to impact argyrin B binding at each site are also highlighted as relevant to substrate specificity. Additional key residues for argyrin B binding also often neighbour those indicated as important to specificity pockets. Other structure based drug design investigations have utilised similar information to achieve β5i selectivity (Groll *et al.*, 2010; Beck *et al.*, 2015; Sosič *et al.*, 2016).

The significantly different, although consistent positioning of argyrin B in each subunit, raises the possibility of an improved and selective fit at β5i. Making changes to the thiazole ring could strongly lower β5c interactions, without affecting β5i since there are few noteworthy interactions currently observed in this location. Whilst changes around the Trp rings may improve the interactions with β5i.

4.2.2.4 Summary of argyrin B and IP selective docking

Molecular docking simulations provided scoring values that matched argyrin B purified assay trends and suggested conformational poses, including interactions that may contribute towards the free energy binding values.

From control tests with TMC-95A (supplementary table 15), 250 runs produced a single best-fit conformation (-9.58 Kcal/mol) predicted at a far greater affinity than all others (starting from -8.57 Kcal/mol). This most favourable conformation of TMC95-A (green in figure 23) was predicted different to the crystal structure pose (yellow) although the single occurrence suggests a possible anomalous result. It is therefore important to also analyse repeats and distinct conformational clusters. With TMC-95A docking results this revealed the most commonly occurring distinct conformational cluster of high scoring poses (example shown in red, figure 23). These do indeed closely match crystallographic data (yellow, figure 23). This highlights a good potential accuracy of

the docking simulations, shown to commonly identify and rank highly the correct pose. However also stresses the care required in analysis to investigate beyond simply the top free energy of binding estimations.

This procedure was applied with argyrin B results show in figure 22 where the frequency of best scoring conformations were analysed. In argyrin B tests (figure 22), no top scoring conformations were isolated occurrences and most also identified within the main distinct conformational cluster (at RMSD of 2 Å), enhancing validity of findings towards matching the true crystallographic pose.

Nonetheless, inconsistencies are observed from docking simulations, as reported in literature (Chen, 2015; Koroleva *et al.*, 2015). To evaluate accuracy of docking software, a root-mean-square deviation (RMSD) between docked and crystallographic poses should be $\leq$2 Å (Guedes *et al.*, 2016). Tests with argyrin B showed low standard deviation from binding energies and relatively high consistency of conformations (figure 22), whilst control tests of TMC-95A reveal predominantly close predictions to crystallographic data.

To supplement figure 33, a summary of differences between active sites that specifically lead to argyrin B preference towards the IP active site, are tabulated in table 20. Many of these are in common, or indirectly impact those deemed important for active site specificity from literature. Whilst others provide novel insight towards methods of achieving IP selectivity. There are some noticeable comparisons that can be drawn between argyrin B docking results at CP sites and similar aforementioned studies by Loizidou and Zeinalipour-Yazdi (2014) and Stauch *et al.*, (2010), despite their use of yeast and humanised models as well as different argyrin analogues (structures shown in figure 12). However there are currently no literature reports available of argyrin analogue docking at the IP.

Table 20 – Key CP and IP structural differences influencing argyrin B interactions - Summary of CP to IP substitutions influencing conformation of argyrin B and possible IP site selectivity.

Residue			Active Site Significance		
Position	β1c	β1i	β1c	β1i	Argyrin B interactions at IP
45	Arg	Leu	Bulky, basic S1	Hydrophobic S1	Altered orientation of molecule
20	Thr	Val	Polar S1	Hydrophobic S1	Unique β1i interactions
31	Thr	Phe	Polar S1	Hydrophobic S1	β1i favours hydrophobic regions of molecule
52	Thr	Ala	Polar S1	Hydrophobic S1	β1i favours hydrophobic regions of molecule
21	Thr	Ser	Structure around S1	Structure around S1	Hydrogen bonding at β1i
97	Gly	His	Small residue	Larger, basic residue	Trp1 interactions at β1i and orientation for Ser46 hydrogen bonding
95	Met	Ser	Hydrophobic	Polar	Trp2 interactions at β1i
114	Val	-	Hydrophobic	Gap - Spatial arrangement	Thiazole proximity closer to Thr1
129	Gly	Ser	Neutral	Polar	Hydrogen bonding in β1i

Residue			Active Site Significance		
Position	β2c	β2i	β2c	β2i	Argyrin B interactions at IP
22	Glu	Asn	Acidic	Polar	Enhance polar region for Thr21 binding at β2i
23	Gly	Asp	Small, neutral	Larger, acidic	Architectural changes to enhance Thr21 binding at β2i
24	Met	Ser	Hydrophobic	Polar	Possible inhibitor stabilisation
48	Thr	Val	Polar S2	Hydrophobic S2	Different orientation
97	Ala	Ser	Hydrophobic	Polar	Trp1 interactions and H bonds in β2i
114	Tyr	His	Hydrophobic	Basic	Alter architecture to enhance β2c P115 and A95 interactions

Residue			Active Site Significance		
Position	β5c	β5i	β5c	β5i	Argyrin B interactions at IP
46	Ala	Ser	Small, hydrophobic	Polar	Double hydrogen bonds at β5i
48	Gly	Cys	Small, hydrophobic S2	Polar S2	Nearby Trp2 interactions at β5i
27	Ala	Ser	Small, hydrophobic S2	Polar S2	Nearby Trp2 interactions at β5i
53	Ser	Glu	Polar	Large, polar	Effects Met45 configuration to enlarge S1 and bind Trp1 in β5i
30	Thr	Arg	Polar	Basic	Structural changes to enhance Val31 and Lys33 β5i interactions around S1
32	Lys	Asn	Basic	Polar	Structural changes to enhance Val31 and Lys33 β5i interactions around S1
21	Thr	Ser	Polar	Smaller, polar	H bonding at β5i

S1 pocket changes are of particular significance in selectivity and activity evidenced by marizomib that achieves potent inhibition by occupying solely S1 (Groll *et al.*, 2006b), yet additional pocket interactions can further enhance potency and selectivity. In fact, non-covalent S3 interactions alone have also shown strong inhibitory properties (Beck *et al.*, 2015). Due to the covalent mechanism of many existing proteasome inhibitors, it is easy to characterise P chains that occupy S1-4 pockets. However, in the case of non-covalent inhibition with greater freedom to interact at different orientations, the S pockets may not always align with the same side groups of the inhibitor, incurring additional challenges to optimisation designs. Despite this, proposed analogues of argyrin A and F by Loizidou and Zeinalipour-Yazdi (2014) exhibited remarkably enhanced affinity for active sites in docking simulations, proving the potential for applying similar techniques to increase selectivity. Therefore with the selectivity of argyrin B towards the IP shown in kinetic assays, combined with understanding of predicted interactions from docking, the IP selectivity could be further enhanced.

Although molecular docking simulations are widely used and provide valuable insight into drug interactions, it is important to back up findings with additional evidence. Secondary docking using FRED tested different approaches to predict inhibitor interactions. These revealed similar findings that help to validate AutoDock predictions (table 15-17). Only β5i raised some concern of AutoDock missing additional interactions. One must appreciate the limitations of computational docking and conclude results appropriately.

4.2.3 Molecular docking software

Molecular docking is becoming an increasingly utilised research tool, with numerous software programmes now available. In table 21, details of AutoDock and FRED docking used in this investigation are summarised alongside 5 of the most commonly published alternative software packages, in recent years. More exhaustive programme reviews can be found elsewhere (Sousa *et al.*, 2013; Chen, 2015).

Table 21 – Overview of most common molecular docking software programmes - Summary of docking software used in this investigation, alongside the next most common platforms. Basic parameters are outlined and all software enable ligand flexibility.

Docking Software	% of Docking Publications 1990-2013	Conformational Search Algorithm	Protein Flexibility	Main Scoring Functions	Reference
AutoDock	25.9	Stochastic	Protein side chains flexible	AutoDock (empirical, force-field based)	(Morris *et al.*, 2009)
GOLD	16.7	Stochastic	Partial flexibility	GoldScore (force-field based) ChemScore (empirical)	(Verdonk *et al.*, 2003)
Glide	14.9	Systematic	Flexible	GlideScore (empirical)	(Friesner *et al.*, 2004)
FlexX	9.6	Systematic	Flexible	FlexX (empirical)	(Kramer *et al.*, 1999)
Surflex-Dock	4.2	Systematic	Flexible	Hammerhead (empirical)	(Spitzer and Jain, 2012)
FITTED	3.3	Systematic	Flexible	Rank Score 5 (force-field based)	(Corbeil *et al.*, 2007)
FRED	n/a	Systematic	Rigid	ChemGauss4 (empirical)	(McGann, 2012)

AutoDock, the most commonly published docking software (from 1990-2013), uses a genetic algorithm with stochastic conformational search whereby a random number generator creates a new location, orientation and torsion values for each pose. This means that results differ between repeats, but a sufficient coverage of the active site can be achieved (Morris *et al.*, 2009). AutoDock estimates free binding energy using equation 6, in which intermolecular energy encompasses Van der Waals forces, hydrogen bonding, desolvation and electrostatic energies. A lower Gibbs free energy represents increased stability of the complex, ranked by empirical or force-field based scoring functions. Docking is performed with a flexible ligand and flexible side chains of the macromolecule; evidently favourable parameters across major docking software (table 21). Although in AutoDock 4.2 most bonds can be made rotatable or non-rotatable, a limitation is that as standard, cyclic portions of the ligand are treated as rigid to avoid distorted structures (Morris *et al.*, 2009). In this investigation, argyrin B was energetically optimised using Avogadro (Hanwell *et al.*, 2012), although additional steps can be undertaken to manage the flexibility of rings, then applied to AutoDock (Forli *et al.*, 2007).

Equation 6 – AutoDock free binding energy calculation – Equation used to derive estimated free enrgy of binding from AutoDock molecular modelling simulations. All components in Kcal/mol units.

$$\textit{Estimated free energy of binding} = \textit{Intermolecular energy} + \textit{Final total internal energy} + \textit{Torsional free energy} - \textit{Unbound systems energy}$$

Literature analysis of multiple docking software revealed striking inconsistencies that highlight the value of repeat testing on additional software (McGann, 2011, 2012). FRED was chosen as an alternative method to provide useful comparison and validation of results in this experiment due to different parameters used. FRED docking utilises a systematic, exhaustive search where gradual variations of conformation are applied; effective for most likely binding modes in localised spaces, although less effective at global minimum (Ferreira *et al.*, 2015). FRED docks with multiple ligand conformers produced by OMEGA that identifies and enumerates rotatable bonds and flexible rings, to essentially mimic flexible ligand docking. Finally, top poses undergo systematic solid body optimisation through local exhaustive searches of rotational and translational degrees of freedom (McGann, 2011).

Alternative software generally follows genetic algorithms and systematic conformational searches. Empirical scoring functions count favourable interactions between 2 molecules, including details such as hydrophobic effect. This method is generally preferred to force-field based scoring functions that add intermolecular and electrostatic interactions between all atoms of the ligand-protein complex (Wang *et al.*, 2002). Alternatively, knowledge-based scoring utilises existing molecule-protein complex crystallisation data to inform favourable binding affinities (McGann, 2012) and would prove valuable tools for optimisation design where ligand-protein crystallisation data is available. Furthermore, the benefit of fully flexible macromolecules of alternative software is a useful function that would allow changes in protein conformation and pocket size upon ligand binding. These induced-fit docking methods have shown an impressive improvement in RMSD values to crystal structures, compared to rigid-receptor docking (Sherman *et al.*, 2006).

Choice of docking software is difficult and highly dependent on experimental aims. Aforementioned docking software each have particular benefits suited for certain needs, for example the ability of GOLD and FITTED to test pharmacophore orientated covalent docking (Verdonk *et al.*, 2003; Corbeil *et al.*, 2007), that would be useful for the majority of existing proteasome inhibitors. To test a broad active site area for non-covalent interactions, AutoDock is appropriate software that has been commonly used for proteasome inhibition investigations (Yang *et al.*, 2008; Kanwar *et al.*, 2010; Loizidou and Zeinalipour-Yazdi, 2014).

It is commonly reported that although predicted ligand binding conformations may correspond with crystallographic data, proving successful docking methods, the estimated energies are often in disagreement between different software and laboratory data (Chen, 2015). This is largely due to a lack of physical data in the model to determine entropic contributions accurately, as well as neglecting solvent considerations that detriment desolvation predictions (later discussed in 4.4.1)

(Ferreira *et al.*, 2015). Indeed, force-field and empirical based scoring functions of AutoDock are liable to these limitations and as such, should be reported with laboratory data.

Advancements from molecular docking are molecular dynamics (MD) simulations with systems such as GROMACS (Hess *et al.*, 2008) and Desmond (Shivakumar *et al.*, 2010) including force fields such as CHARMM (Brooks *et al.*, 1983) and AMBER (Cornell *et al.*, 1995). Ligand parameter and topology identification for use in MD programmes can be generated using software such as SwissParam (Zoete *et al.*, 2011) and LigParGen (Dodda *et al.*, 2017). These methods are difficult to research due to heavy demands on computational power, but would test and reveal any physical movements of molecules over time, upon ligand binding (Ferreira *et al.*, 2015). Full flexibility of the target and monitoring changes upon ligand interaction may reveal cryptic pockets and dynamic binding conformations. 3 Dynamic states of the 26S proteasome are known to allow recognition and entry into the 20S core where it is possible a ligand causes further proteasome flexibility (Huang *et al.*, 2016).

4.3 Current IP inhibitor status

4.3.1 Advances in structure-based drug design

The importance of suitable proteasome structures is highlighted by species variations. Geurink *et al.*, (2013) show the difference between use of yeast or mammalian structures that reveal how single residue changes can have a significant effect upon structure related activity studies. For example, oprozomib and PR-924 show 20 and 130-250 fold selectivity to human β5i over β5c (Parlati *et al.*, 2009), with the latter on account of worse β5c affinity rather than greater β5i. However, this difference is not observed between mouse ratio of β5i and β5c affinity where oprozomib selectivity remains 20-fold yet PR-924 is reduced to 16-fold (Huber *et al.*, 2016). It is therefore important to use models that map as closely as possible to the human CP and IP. There are examples of successful use of homology models and molecular docking approaches, similar to those described in this investigation, to test proteasome inhibitors (Loizidou and Zeinalipour-Yazdi, 2014; Guedes *et al.*, 2016). These are required because crystallographic structure determination involves challenges such as: overcoming high mobility, losing protein fractions and combination with other structures (Ferreira *et al.*, 2015). Therefore, human CP structures were only recently elucidated (Harshbarger *et al.*, 2015) whilst human IP structural data is not currently available.

Crystal structures of compound and proteasome help to fully elucidate binding interactions and are considered the gold standard of testing inhibitor interactions. Therefore, to overcome species differences and crystallisation limitations, sophisticated chimeric humanised yeast proteasomes have recently been developed (Huber *et al.*, 2016). Human β5i fragment chain 1-138 and neighbouring β6 subunit residues that can be involved in ligand interactions, were incorporated into yeast proteasomes then crystallised to create a structural model. X-ray analysis with inhibitors and mutant variants revealed the cause of species difference between PR-924 selectivity. The Met31 in murine β5i increases the P1 size and causes steric hindrance for inhibitors such as PR-924 endowed with bulky N-caps, although those with smaller groups including oprozomib can be accommodated. Whilst substituted Val31 in yeast humanised chimeric β5i allows a kinked binding mode with increased potency. These findings further validate structure-based design and introduce a yeast proteasome with incorporated human β5i subunit, as a powerful tool for advanced inhibitor studies (Huber *et al.*, 2016). Similar methods have also been used on a smaller scale of single mutations on yeast proteasomes then crystallised to reveal residue orientation (Duibella *et al.*, 2015). However, despite credibility for crystallisation studies, these chimeric proteasomes are not advised as an alternative for enzymatic assays as β5 activity is significantly reduced by growth irregularities and temperature sensitisation (Huber *et al.*, 2016).

Supplementary figure 17 shows an alignment of yeast proteasome with human β5i (in red) (Huber *et al.*, 2016, Cui *et al.*, 2017) to humanised β5i homology model (green) created in this investigation, with differences in blue and β5i subunit chains in magenta. Significant changes within the β5 chains are shown at Thr1 and Gly188 with the remainder closely matched. Gly188 is positioned towards the back of the site and not involved in ligand binding, however slightly differing positioning of the catalytic Thr1 is of interest. As expected, some changes are observed in neighbouring chains within the 28 Å cut, as the yeast proteasome differs from mouse IP. This shows the humanised β5i homology model as suitable due to close resemblance of human β5i structural data; however the minor difference in Thr1 position may influence docking. It is expected that β1i and β2i humanised models would achieve comparable findings to structural data, if available.

Since the human 20s CP was first crystallised to a resolution of 2.6 Å (Harshbarger *et al.*, 2015), advances in structural elucidation have revealed further, novel data. A cryo-electron microscopy structure of the human 26S CP at 3.5 Å reveals greater conformational dynamics for understanding substrate processing (Huang *et al.*, 2016). Additionally, a series of extra production, purification and crystallisation steps accompanied by technological advances have led to a 1.8 Å human 20S CP structure with clear electron densities (Schrader *et al.*, 2016). Novel co-

crystallisation studies of the CP in complex with inhibitors revealed differences in binding mechanisms to those originally proposed and further characterisation of the β5c active site (Schrader *et al.*, 2016). Specifically, the previous epoxyketone mechanism thought to form a 1-4 morpholino adduct formation has been shown to in fact produce a seven-membered ring adduct between the inhibitor and threonine residue (Schrader *et al.*, 2016). With this newfound information, new classes of boronic acid or epoxyketone inhibitors may be developed based upon formation of a seven-membered adduct, allowing for example, different positioning of the leaving group. Additionally, electron diffraction techniques have recently been implemented to reveal hydrogen atom localisation in organic material, a current limitation to x-ray diffraction crystallisation data (Palatinus *et al.*, 2017). Overall, these higher resolution 3-dimensional structures allow more accurate details on the topology of the macromolecule sub pockets, clefts and cavities, whilst also providing electrostatic properties. Therefore, where possible, the highest resolution data should therefore be used in molecular docking studies.

A crystallographic screening procedure approach has recently been utilised with yeast CP on a set of compounds, providing rapid elucidation of inhibitor-proteasome complex structural data (Beck *et al.*, 2015). This is a powerful approach allowing detailed investigations on a wider variety of compounds; effectively used to identify agents with a novel mode of action.

These novel advances and approaches have rapidly increased the development of proteasome inhibitors, particularly in understanding and achieving selectivity.

4.3.2 Immunoproteasome selective inhibitors

Using approaches such as high throughput screening and synthetic modifications of existing inhibitors based upon structural data, an increasing number of IP selective inhibitors are being discovered. Compounds to date, with proposed IP selectivity, quantified through IC_{50} values, are summarised in table 22.

Table 22 - IP selective inhibitors determined by IC_{50} values at each active site – Inhibitors are grouped by class of binding mode and a brief outline of methods is included. * represents K_{inact}/K_i ($M^{-1}s^{-1}$), a preferred measure for irreversible inhibitors.

Class	Compound	IC_{50} (μM) β5c	IC_{50} (μM) β5i	Ratio of IP Selectivity	Method
Peptidyl aldehyde	IPSI-001 (calpeptin)	2.9	0.022	131	Purified assays (Parlati *et al.*, 2009)
		Ki 105	Ki 1.03	102	Z-GGF-pAB purified assays (Kuhn *et al.*, 2009)
Epoxyketone	ONX 0914 (PR-957)	0.236	0.028	8	Suc-LLVY-AMC purified assays (Muchamuel *et al.*,, 2009)
		0.16	0.015	11	(Sosič *et al.*, 2016)
		0.513	0.057	9	(Duibella *et al.*, 2014)
		0.487	0.024	20	(Huber *et al.*, 2016)
		0.054	0.0057	9.5	Raji cell lysates, activity-based protein profiling with BODYIP-NC005 probe (de Bruin *et al.*, 2014)
	From ONX 0914:				
	Lu-005i (8)	0.29	0.0066	44	Raji cell lysates. Activity based protein profiling with BODYIP-NC005 probe (de Bruin *et al.*, 2014)
	Lu-025i (11)	1.9	0.036	52	
	Lu-045i (15)	0.83	0.032	26	
	Peptido sulfonyl fluoride compound 3	3.927	0.139	28	Suc-LLVY-AMC purified assays (Duibella *et al.*, 2014)
	PR-924	13.02	0.051	255	Suc-LLVY-AMC purified assays (Huber *et al.*, 2016)
	From PR-924:				
	Lu-015i (9)	4.6	0.0083	554	Raji cell lysates. Activity based protein profiling with BODYIP-NC005 probe (de Bruin *et al.*, 2014)
	Lu-035i (13)	5.5	0.011	500	
	Lu-055i (25)	25	0.053	471	
Chloroacetamide	Carfilzomib derived decarboxylated peptide with amide P3 sidechain	>100	0.64	>150	Purified assays (Dubiella *et al.*, 2015)
Oxathiozole	HT2004	0.23*	1093*	4750	Suc-LLVY-AMC purified assays (Fan *et al.*, 2014)
	HT2106	0.7*	151*	215	

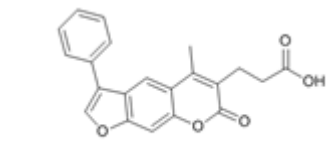
	HT2210	3.2*	83*	26	
	Compound 42	1.07	0.013	82	Suc-LLVY purified assays. (Sosič *et al.*, 2016)
Non-peptide, reversible	Psoralene based compound 3	Ki 172.2	Ki 1.6	108	Suc-LLVY purified assays. (Sosič *et al.*, 2016)
Non-covalent	Capped dipeptide inhibitor Compound 4 Compound 5	470 340	41 27	11 13	Ac-WLA-AMC, Ac-ANW-AMC purified assays (Blackburn *et al.*, 2010)
Non-covalent, non-peptidic	Ro19	28.85	0.42	69	Proteasome-Glo™ Suc-LLVY-aminoluciferin substrate (Cui *et al.*, 2017)

Class	Compound	IC_{50} (μM)		Ratio of IP Selectivity	Method
		β1c	β1i		
Peptidyl aldehydes	IPSI-001 (calpeptin) Peptidyl aldehydes	Ki 239	Ki 1.45	165	BrAAP assay (Kuhn *et al.*, 2009)
Epoxyketone	UK-101	15	0.104	144	Raji cell lysates - Activity based protein profiling with BODYIP-NC001 probe (Ho *et al.*, 2007; de Bruin *et al.*, 2014)
	Lu-001i	24	0.095	252	Raji cell lysates - Activity based protein profiling with BODYIP-NC001 probe (de Bruin *et al.*, 2014)
		47	0.13	360	RPMI-8226 cells (de Bruin *et al.*, 2014)

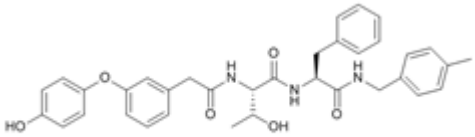
Class	Compound	IC_{50} (μM)		Ratio of IP Selectivity	Method
		β2c	β2i		
Non covalent bisbenzimidazoles	Monomeric	2	0.024	83	Boc-LRR-AMC purified assays (Koroleva *et al.*, 2015)
	Dimeric	4.1	0.024	171	

There are currently a higher number of β5i selective compounds discovered on account of aforementioned structural differences and research greater focus on the β5 site, often considered the most influential. In addition, despite an increasing number of non-covalent inhibitors, summarised in table 3, few of these display IP selectivity. Therefore, the β1i selectivity and non-covalent interactions of argyrin B are interesting findings that collectively are unique to existing compounds. Figure 34 displays chemical structures of non-covalent IP selective inhibitors and β1i selective inhibitors, of which all are covalent.

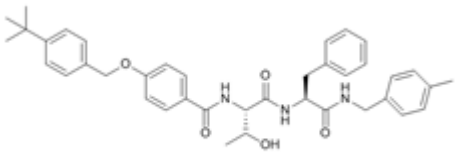
Non-covalent β5i selective inhibitors



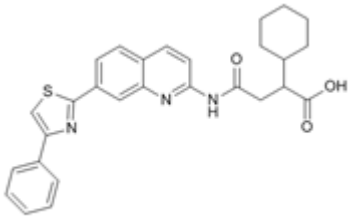
A) Psoralene based compound 3



B) Capped dipeptide inhibitor 4

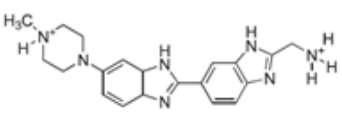


C) Capped dipeptide inhibitor 5

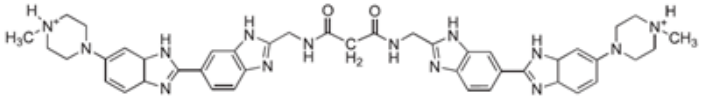


D) Ro19

Non-covalent β2i selective inhibitors

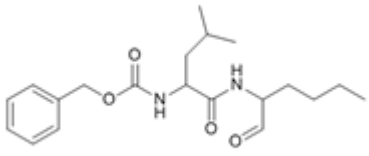


H) Monomeric bisbenzimidazole

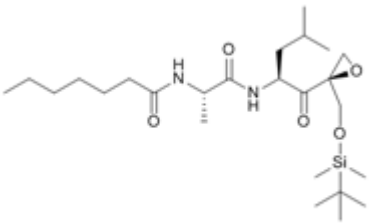


I) Dimeric bisbenzimidazole

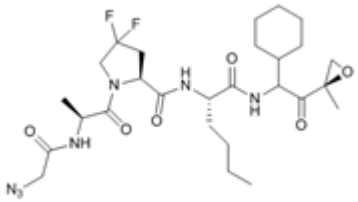
Covalent β1i selective inhibitors



E) IPSI-001



F) UK-101



G) Lu-001i

Figure 34 – Structures of IP selective inhibitors – β5i selective and non-covalent: A) (Sosič *et al.*, 2016), B) (Blackburn *et al.*, 2010), C) (Blackburn *et al.*, 2010), D) Ro19 (Cui *et al.*, 2017). β2i selective and non-covalent: H) and I) (Koroleva et al., 2015). β1i selective, covalent: E) (Kuhn *et al.*, 2009), F) and G) (de Bruin *et al.*, 2014).

Purified assays, as used in this investigation, are popular methods for kinetic studies with proteasome inhibitors. However, variations can be observed between repeat tests of different research groups, as shown with PR-957 in table 22. Approximately 3-4 fold differences can be observed between individual IC_{50} values; however the measure of IP selectivity ratio is slightly more consistent. Indeed the inconsistencies of enzyme assays have been previously reported; largely due to incubation times, particularly with covalent inhibitors and the varied, single substrate concentration of IC_{50} tests. Suggested actions emphasise the value of K_i measurements over IC_{50} (Cisneros *et al.*, 2016a).

Alternative methods are also reported, notably use of raji cell lysates for activity based protein profiling probes (de Bruin *et al.*, 2014). B-cell lymphoma cell lines are incubated with inhibitor concentrations and following cell lysis, specific probes added to bind active sites. SDS-PAGE (polyacrylamide gel electrophoresis) separation then allows fluorescent densitometry of bands for quantification (de Bruin *et al.*, 2014). This method has advantages of using a cellular environment, however arguably not as sensitive as purified fluorescence assays for ascertaining kinetic profiles. Advances in detection probes and measurements have also revealed potential use for fluorescence polarisation-based assays, as an efficient and accurate inhibitor binding measurement for dissociation constant (K_d) determination (Cisneros *et al.*, 2016b). Further laboratory based methods are previously outlined in section 1.6.1.

Additional argyrin B tests with alternative assay methods would prove useful to investigate the effects in cells and across all active sites together. Further analysis of mechanism using knockout models as described with argyrin A by Nickeleit et al., (2008) would also prove valuable information towards drug development. However, initially strategies to enhance argyrin B potency would be a priority.

Although the potency observed by argyrin B is relatively low in comparison to ranges observed in table 22, understanding the interactions and mode of binding can facilitate further development. Synthetic additions of groups can further enhance potency. Other IP selective inhibition screens and studies have revealed proof-of-principle and leads for further investigation, despite selectivity in only low micromolar ranges (Kuhn *et al.*, 2009). To investigate, free energy perturbation (FEP) calculation simulations are a valuable approach, using aforementioned molecular dynamics packages (section 4.2.3). Able to compute differences in free energy from chemical perturbations of a compound, this can reduce the demand on physical synthesis and facilitate greater variety of testing (Wang *et al.*, 2015).

An argyrin B IP selectivity ratio of 16 (table 8) measures against the capped dipeptide non-covalent β5i inhibitor (structure in figure 34) (Blackburn *et al.*, 2010) in terms of ratio and potency. However compared to β1i inhibitors (Kuhn *et al.*, 2009; de Bruin *et al.*, 2014) and the non-covalent bisbenzimidazole β2i inhibitors (Koroleva *et al.*, 2015), potency and selectivity are relatively low. Nonetheless, predicted interactions that result in the binding preferences observed table 20, may prove valuable information in future development of selective inhibitor drug design.

4.4 Evaluation

4.4.1 Experimental limitations

There remain limitations associated with the results gathered in this investigation, some of which have been previously mentioned.

Whilst K_i values have been estimated these are currently single tests that require further repeats, the value of which is evidenced by variation to IC_{50} values. Where possible the inhibitor concentrations should range from 0, 0.33, 1 & 3 x IC_{50} value, whilst substrate concentrations cover 0.2-5 x K_m (Copeland, 2015). However, due to plate reader detection limits and the presence of additional controls limiting space on a 96-well plate, these were not always experimentally practical. Importance of K_i values is clear (Cisneros *et al.*, 2016a) therefore further tests to confirm and perform statistical analysis on argyrin B K_i values would enhance conclusions.

DMSO solvent was essential to solubilise argyrin B and substrate assay components (table 5); water solubility of argyrin A and F have been characterised at low micromolar levels (Bülow *et al.*, 2010). As compound stocks were stored in DMSO it is possible that the stability is affected by freeze-thaw cycles increasing moisture absorption from the air (Cheng *et al.*, 2003). Where possible aliquots or fresh stocks were prepared, however there may remain a source of error with unrepresentative compound concentrations after prolonged freezing and use. Furthermore, at high concentrations of compound, the respective final DMSO levels remained high and exhibited significant effect upon the rate of reaction (figure 19). Whilst concurrent DMSO control correction at corresponding DMSO concentrations accounted for the impact of solvent, the consequence towards reaction conditions must also be considered. Despite correction, the impact upon reagent concentrations can be significant enough to create a different reaction environment that is no longer appropriate for kinetic studies.

The type of microplate used can be influential, with (Cui *et al.*, 2014) revealing the significant impact of colour and binding surface. In this experiment, standard assay kit component NBS white plates although this is likely to vary amongst other studies.

Whilst at the time of experiments, PDB:4R30 (Harshbarger *et al.*, 2015) was the highest resolution human CP structural data available, there is now greater detail available in the PDB:5LE5, 1.8 Å data (Schrader *et al.*, 2016). This would be a more appropriate source to create active site cuts used in molecular docking. In homology preparation of humanised IP structures, newly substituted residues are predicted orientations at the given site. However, the recent crystallisation of humanised yeast β5i by Huber *et al.*, (2016) provide more solid evidence of residue positioning and side chain orientation. Due to the minor change in Thr1 position observed (supplementary figure 17), this would be considered more appropriate data for β5i docking experiments.

The type of docking software used can include some limitations as previously discussed in 4.2.3, particularly compared to the rapidly emerging molecular dynamic programmes. Inability to test argyrin B with a fully flexible active site target may limit the simulations performed in this investigation.

Furthermore, the absence of water molecules in AutoDock is considered a limitation when compared to the true *in vivo* state. Water molecules can possess a variety of thermodynamic energetic states, including 'unhappy' waters often in lipophilic regions due to preference over the absence of water, yet disfavoured to positioning within the bulk solvent (Mason *et al.*, 2013). Therefore energetically 'unhappy' water may be displaced by ligand binding to augment strength, 'happy' water may be removed in a less favourable binding, or water molecules may be trapped by a ligand. Additionally, new water molecules forming around the ligand-protein complex can enhance entropy and mediate interactions to residues (Beuming *et al.*, 2012). Understanding and including the positioning and energetics of water molecules is therefore valuable with ligand-binding simulations. WaterMAP (Abel *et al.*, 2008) and WaterFLAP (Mason *et al.*, 2013) may be used in conjunction with molecular dynamics software to assess the impact of water and potentially be exploited to enhance selectivity (Beuming *et al.*, 2012; Mason *et al.*, 2013).

Results concluded from purified assay and docking experiments can be limited. Whilst inhibition of an active site can be observed, this may not translate in cellular or *in vivo* conditions. In more complex environments, issues such as mixed active sites and compensatory proteolytic activity may reduce or silence inhibition previously observed (Kisselev *et al.*, 2006). It is also argued that proteasome variation (figure 7), post-translational modifications and other interacting proteins

may cause differences observed in active site specificities (Geurink *et al.*, 2013). Therefore, results are limited to impacts on a simplified model and only proposed or indicative cellular effects and therapeutic potential.

4.4.2 Future Developments

Whilst this investigation indicates argyrin B selectivity towards the IP, further repeats are required to validate K_i values. In addition, determination of K_d values would be valuable to define reversibility strength and lasting action of the inhibitor (Cisneros *et al.*, 2016b). When sufficient kinetic data is collected, it is important to consider the way in which inhibitor data is reported. It is commonly suggested to follow guidelines outlined as standards for reporting enzyme data (STRENDA) to sufficiently report experimental detail and functional data from enzyme kinetic experiments (Tiptona *et al.*, 2014)(full guidelines provided at http://www.beilstein-institut.de/en/projects/strenda/guidelines should be used when possible). Within a field containing frequent high throughput screening and early stage compound characterisation, these details of kinetic data using rigorous methods are often overlooked, despite the value of such information.

Argyrin B has shown promising early signs of IP selectivity and information has been gathered on possible interactions that contribute towards a greater affinity at IP sites. Furthermore, with increased understanding of each active site and proposed binding characteristics, modifications to the inhibitor may enhance potency and or selectivity. It would be logical to first test this computationally through molecular docking which can be quantified through comparison of binding energies. Further to this, upon synthetic development of a new analogue, the difference in K_i values can be indicative of binding affinity that could be further quantified by change in binding energy using equation 7 (Copeland, 2005).

Equation 7 - Gibbs free binding energy - R = ideal gas constant, T = temperature (kelvin), K_i – inhibition constant (moles).

$$\Delta Gbinding = RTln(Ki)$$

Equation 7 can also be modified to calculate the difference in binding energies between 2 inhibitors, by using a ratio of K_i^A/K_i^B. Differences in calculated binding energy can also help to

confirm the type of interactions that are introduced. For example, a binding energy difference of approximately -2.5 kcal/mol would be indicative of an extra hydrogen bond which may be attributed to the structural change of the compound (Copeland, 2005). Combined with computational modelling, this can be a powerful approach to validate predictions and optimise compounds. This approach would be particularly useful with IP studies where co-crystallisation is difficult to achieve. At the CP it would be valuable to co-crystalise argyrin B to confirm binding orientation and interactions. With novel chimeric yeast proteasome with human β5i subunit (Huber *et al.*, 2016), argyrin B may also be crystallised into the human β5i site, as achieved with Ro19 inhibitor by Cui *et al.*, (2017).

Beyond simplified, pharmacokinetic assays, proteasome inhibitors can be measured in whole blood samples and more complex models (Lightcap *et al.*, 2000; Parlati *et al.*, 2009). These experiments can be useful to identify pharmacodynamic properties and molecular mechanisms of the inhibitor. Such experiments have been performed with argyrin A revealing anti-tumoural activity achieved through preventing the breakdown of $p27^{kip1}$ (Nickeleit *et al.*, 2008). As this is a unique mechanism (figure 9) to existing proteasome inhibitor therapeutics (figure 11), there are associated benefits of potentially overcoming resistance in relapsed cases. It would therefore be interesting to investigate whether the argyrin B analogue displays the same mechanism of action. Exposure of the compound in cell lines followed by western blot analysis may be used to test the impact upon key regulatory protein levels. However, there are also difficulties that arise from testing in cell lines. Such as variance in permeability across cell lines and in some inhibitor classes off target effects for example cathepsin protease inhibition, can occur making it difficult to establish the true causality (Geurink *et al.*, 2013).

With the suggestion of non-competitive inhibition of argyrin B, it would be interesting to test argyrin binding at α-subunits. Competition assays with chloroquine can be performed as described by Stauch *et al.*, (2010). Here resonance transferred-NOE signals measured chloroquine that is able to bind in the presence of argyrin A, where chloroquinone would typically bind α-subunits.

Additionally, some protein targets have recently discovered cryptic pockets that reveal upon drug binding. In some cases these allosterically alter enzymatic activity and could provide alternative targeting sites (Bowman and Geissler, 2012). Although not currently identified within literature for proteasomes, advances in molecular dynamic modelling software combined with enhanced computational power, may allow longer simulations and further details of protein flexibility. Future investigations with software such as TRAnsient pockets in proteins (TRAPP) could provide information on changes in a binding pocket's spatial and physiochemical properties, potentially

revealing cryptic sub-pockets (Stank *et al.*, 2017).

Finally, if drug development were to continue as a lead compound, complex tests and considerations are required regarding appropriate solubility, permeability, metabolism and toxicity (Strelow *et al.*, 2012). Basic analysis of argyrin B absorption, distribution, metabolism and excretion, contributing towards drug-likeness of the molecule are summarised in figure 35, calculated using SwissADME (Daina *et al.*, 2017).

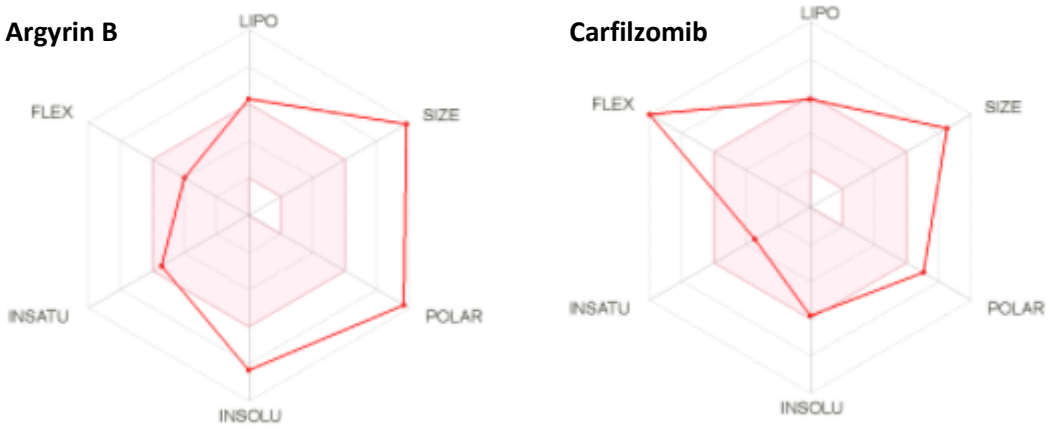


Figure 35 – Drug-likeness properties of argyrin B and carfilzomib – Shaded red zone represents guideline suitable physiochemical properties for oral bioavailability. Measures of lipophilicity (LIPO), size, polarity, insolubility (INSOLU), insaturation (INSATU) and flexibility (FLEX) (Daina *et al.*, 2017).

Figure 35 reveals potential challenges in size, solubility and polarity of argyrin B, in medicinal chemistry. These would particularly cause challenges in drug delivery and absorption. Furthermore, 7 hydrogen bond donors are above the recommended 5 maximum, combined with a molecular weight above 500, cause argyrin B to violate the Lipinski drug-likeness recommendations (Lipinski *et al.*, 2001). For comparison, figure 35 also shows drug development difficulties in polarity, size and flexibility for carfilzomib, yet these have been overcome and FDA approval has been successful.

4.5 Conclusion

In conclusion, the naturally occurring, cyclic peptide argyrin B was tested for inhibition at the CP and IP. Purified enzymatic assays reveal argyrin B IC_{50} values of 8.76 µM at β1i that is 16-fold lower and statistically significant to 146.5 µM at β1c. Meanwhile, low micromolar IC_{50} values are observed at β5 sites, with slight preference of β5i over β5c. The same trend is corroborated by K_i dissociation constants and molecular docking estimated free binding energies. AutoDock and OEDocking FRED simulations predicted best-fit binding poses and key interactions of argyrin B at each CP and IP active site, revealing differences that are associated with contributing towards selectivity. T20V, T31F and T52A substitutions from β1c to β1i appear significant in creating a more hydrophobic β1i S1 pocket that enhances multiple argyrin B interactions. T21S from β1c to β1i allow β1i hydrogen bonding with argyrin B and facilitate orientation for multiple Trp moiety interactions. At β5c a small, hydrophobic S2 pocket is replaced with more polar characteristics at β5i that promote argyrin B interactions. In addition, V31 and K33 are identified as important residues surrounding the β5i S1 pocket.

These findings reveal argyrin B as a novel β1i over β1c selective proteasome inhibitor with reversible, non-covalent binding mode. Select structural differences are suggested to contribute towards selectivity and may be beneficial information towards future design of IP selective compounds. The non-covalent mechanism and ability to target the IP over CP is associated with lower toxicity of proteasome inhibitors that exhibit great therapeutic potential against a variety of diseases.

Reference List

Abel, R., Young, T., Farid, R., Berne, B.J. and Friesner, R.A. (2008) 'Role of the active-site solvent in the thermodynamics of factor Xa ligand binding', *J Am Chem Soc*, *130*(9), pp.2817-2831.

Azakir, B.A., Erne, B., Di Fulvio, S., Stirnimann, G. and Sinnreich, M. (2014) 'Proteasome inhibitors increase missense mutated dysferlin in patients with muscular dystrophy', *Sci Transl Med*, 6(250), e: 250ra112.

Basler, M., Dajee, M., Moll, C., Groettrup, M. and Kirk, C.J. (2010) 'Prevention of Experimental Colitis by a Selective Inhibitor of the Immunoproteasome', *J Immunol.*, 185(1), pp. 634-41.

Basler, M., Mundt, S., Muchamuel, T., Moll, C., Jiang, J., Groettrup, M. and Kirk, C.J. (2014) 'Inhibition of the immunoproteasome ameliorates experimental autoimmune encephalomyelitis', *EMBO Mol Med*, 6(2), pp. 226-38.

Beck, P., Lansdell, T.A., Hewlett, N.M., Tepe, J.J. and Groll M. (2015a) 'Indolo-phakellins as β5-specific noncovalent proteasome inhibitors', *Angew Chem Int Ed Engl.*, 54(9), pp. 2830-3, doi: 10.1002/anie.201410168.

Beck, P., Reboud-Ravaux, M. and Groll, M. (2015b) 'Identification of a β1/β2-specific sulfonamide proteasome ligand by crystallographic screening', *Angew Chem Int Ed Engl.*, 54(38), pp. 11275-8, doi: 10.1002/anie.201505054.

Bellavista, E,. Andreoli, F., Parenti, M.D., Martucci, M., Santoro, A., Salvioli, S., Capri, M., Baruzzi, A., Del Rio, A., Franceschi, C. and Mishto, M. (2013) 'Immunoproteasome in cancer and neuropathologies: a new therapeutic target?', *Curr Pharm Des.*, 19(4), pp. 702-18.

Berkers, C.R., Verdoes, M., Lichtman, E., Fiebiger, E., Kessler, B.M., Anderson, K.C., Ploegh, H.L., Ovaa, H. and Galardy, P.J. (2005) 'Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib', *Nature*, 2(5), pp. 357-62, doi:10.1038/NMETH759.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, N., Shindyalov, I.N. and Bourne, P.E. (2000) 'The Protein Data Bank', *Nucleic Acids Research*, 28, pp. 235-42.

Beuming, T., Che, Y., Abel, R., Kim, B., Shanmugasundaram, V. and Sherman, W. (2012) 'Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization'. *Proteins*, *80*(3), pp.871-883.

Bisswanger, H. (2014) 'Enzyme Assays', *Persp. in Science*, 1(1), pp. 41-55, doi: 10.1016/j.pisc.2014.02.005.

Blackburn, C., Gigstad, K.M., Hales, P., Garcia, K., Jones, M., Bruzzese, F.J., Barrett, C., Liu, J.X., Soucy, T.A., Sappal, D.S., Bump, N., Olhava, E.J., Fleming, P., Dick, L.R., Tsu, C., Sintchak, M.D. and Blank, J.L. (2010) 'Characterization of a new series of non-covalent proteasome inhibitors with exquisite potency and selectivity for the 20S beta5-subunit', *Biochem J.*, 430(3), pp. 461-76, doi: 10.1042/BJ20100383.

Bowman, G.R. and Geissler, P.L. (2012) 'Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites', *PNAS*, 109(29), pp. 11681-86, doi: 10.1073/pnas.1209309109.

Brooks, A.D., Jacobsen, K.M., Li, W., Shanker, A. and Sayers, T.J. (2010) 'Bortezomib sensitizes human renal cell carcinomas to TRAIL apoptosis through increased activation of caspase-8 in the death-inducing signaling complex', *Mol Cancer Res.*, 8(5), pp. 729-38, doi: 10.1158/1541-7786.MCR-10-0022.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.A. and Karplus, M. (1983) 'CHARMM: A program for macromolecular energy, minimization, and dynamics calculations', *J. Comp. Chem*, 4(2), pp. 187–217.

Bülow, L., Nickeleit, I., Girbig, A.K., Brodmann, T., Rentsch, A., Eggert, U., Sasse, F., Steinmetz, H., Frank, R., Carlomagno, T., Malek, N.P. and Kalesse, M. (2010) 'Synthesis and biological characterization of argyrin F', *ChemMedChem.*, 5(6), pp. 832-6.

Chen, C.N., Graber, T.G., Bratten, W.M., Ferrington, D.A. and Thompson, L.V. (2014) 'Immunoproteasome in animal models of Duchenne muscular dystrophy', *J Muscle Res Cell Motil*, 35(2), pp. 191-201.

Chen, S., Kammerl, I.E., Vosyka, O., Baumann, T., Yu, Y., Wu, Y., Irmler, M., Overkleeft, H.S., Beckers, J., Eickelberg, O., Meiners, S. and Stoeger, T. (2016), 'Immunoproteasome dysfunction augments alternative polarization of alveolar macrophages', *Cell Death and Differentiation*, 23, pp. 1026–37.

Chen, X., Bui, K.C., Barat, S., Thi Nguyen, M.L., Bozko, P., Sipos, B., Kalesse, M., Malek, N.P. and Plentz, R.R. (2017) 'Therapeutic effects of Argyrin F in pancreatic adenocarcinoma', *Cancer Lett.*, 399, pp. 20-8, doi: 10.1016/j.canlet.2017.04.003.

Chen, Y.C. (2015) 'Beware of docking!', *Trends Pharmacol Sci.*, 36(2), pp. 78-95.

Cheng, X., Hochlowski, J., Tang, H., Hepp, D., Beckner, C., Kantor, S. and Schmitt, R. (2003) 'Studies on repository compound stability in DMSO under various conditions', *J Biomol Screen*, 8(3), pp. 292-304.

Cheng, Y. and Prusoff, W.H. (1973) 'Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction', *Biochem Pharmacol*, 22(23), pp. 3099-108.

Cisneros, J.A., Robertson, M.J., Valhondo, M. and Jorgensen, W.L. (2016a) 'Irregularities in enzyme assays: The case of macrophage migration inhibitory factor', *Bioorganic & med chem lett*, *26*(12), pp.2764-2767.

Cisneros, J.A., Robertson, M.J., Valhondo, M. and Jorgensen, W.L. (2016b) 'A Fluorescence Polarization Assay for Binding to Macrophage Migration Inhibitory Factor and Crystal Structures for Complexes of Two Potent Inhibitors', *J Am Chem Soc*, *138*(27), pp.8630-8638.

Citovsky, V., Zaltsman, A., Kozlovsky, S.V., Gafni, Y. and Krichevsky, A. (2009) 'Proteasomal degradation in plant-pathogen interactions', *Semin Cell Dev Biol.*, 20(9), pp. 1048-54.

Clerc, J., Groll, M., Illich, D.J., Bachmann, A.S., Huber, R., Schellenberg, B., Dudler, R. and Kaiser, M. (2009) 'Synthetic and structural studies on syringolin A and B reveal critical determinants of selectivity and potency of proteasome inhibition', *Proc Natl Acad Sci USA*, 106(16), pp. 6507-12, doi: 10.1073/pnas.0901982106.

Copeland, R.A. (2005) 'Evaluation of Enzyme Inhibitors in Drug Discovery - A guide for Medicinal Chemists and Pharmacologists', *Wiley*, Vol. 46.

Corbeil, C.R., Englebienne, P. and Moitessier, N. (2007) 'Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0', *J Chem Inf Model.*, 47(2), pp. 435-49.

Cornell, W.D. (1995) 'A second generation force field for the simulation of proteins, nucleic acids, and organic molecules' *J. Am. Chem. Soc*, 117, pp. 5179–97.

Cornish-Bowden, A . (2014) 'Analysis and interpretation of enzyme kinetic data', *Science*, 1(1), pp. 121-25, doi: 10.1016/j.pisc.2014.02.010.

Cortés, A., Cascante, M., Cárdenas, M.L. and Cornish-Bowden, A. (2001) 'Relationships between inhibition constants, inhibitor concentrations for 50% inhibition and types of inhibition: new ways of analysing data', *Biochem J*, 357(1), pp. 263-8.

Cuervo, A.M. and Dice, J.F. (1996) 'A receptor for the selective uptake and degradation of proteins by lysosomes', *Science*, 273(5274), pp: 501-3.

Cui, H., Baur, R., Le Chapelain, C., Dubiella, C., Heinemeyer, W., Huber, E.M. AND Groll, M. (2017) 'Structural Elucidation of a Nonpeptidic Inhibitor Specific for the Human Immunoproteasome', *Chembiochem*, 18(6), pp. 523-6, doi: 10.1002/cbic.201700021.

Cui, Z., Gilda, J.E. and Gomes, A.V. (2014) 'Crude and purified proteasome activity assays are affected by type of microplate', *Anal. Biochem*, 446, pp. 44-52, doi: 10.1016/j.ab.2013.10.018.

Daina, A., Michielin, O. and Zoete, V. (2017) 'SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules', *Sci Rep*, 7(42717), doi: 10.1038/srep42717.

Dantuma, N.P., Lindsten, K., Glas, R., Jellne, M. and Masucci, M.G. (2000) 'Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells', *Nat Biotechnol.*, 18(5), pp. 538-43.

de Bruin, G., Huber, E.M., Xin, B.T., van Rooden, E.J., Al-Ayed, K., Kim, K.B., Kisselev, A.F., Driessen, C., van der Stelt, M., van der Marel, G.A., Groll, M., and Overkleeft, H.S. (2014) 'Structure-based design of β1i or β5i specific inhibitors of human immunoproteasomes', *J Med Chem.*, 57(14), pp. 6197-209, doi: 10.1021/jm500716s.

Díaz-Hernández, M., Hernández, F., Martín-Aparicio, E., Gómez-Ramos, P., Morán, M.A., Castaño, J.G., Ferrer, I., Avila, J. and Lucas, J.J. (2003) 'Neuronal Induction of the Immunoproteasome in Huntington's Disease', *J Neurosci.*, 23(37), pp. 11653-61.

Dick, T.P., Nussbaum, A.K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanović, S., Wolf, D.H., Huber, R., Rammensee, H.G. and Schild, H. (1998) 'Contribution of Proteasomal β-Subunits to the Cleavage of Peptide Substrates Analyzed with Yeast Mutants', *J Biol Chem.*, 273(40), pp. 25637-46.

Dodda, L.S., Cabeza de Vaca, I., Tirado-Rives, J. and Jorgensen, W.L. (2017) 'LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands', *Nucleic Acids Res*, doi: 10.1093/nar/gkx312.

Dubiella, C., Cui, H., Gersch, M., Brouwer, A.J., Sieber, S.A., Krüger, A., Liskamp, R.M. and Groll, M. (2014) 'Selective inhibition of the immunoproteasome by ligand-induced crosslinking of the active site', *Angew Chem Int Ed Engl.*, 53(44), pp. 11969-73, doi: 10.1002/anie.201406964.

Dubiella, C., Baur, R., Cui, H., Huber, E.M. and Groll, M. (2015) 'Selective Inhibition of the Immunoproteasome by Structure-Based Targeting of a Non-catalytic Cysteine', *Angew Chem Int Ed Engl.*, 54(52), pp. 15888-91, doi: 10.1002/anie.201506631.

Engelhard, V.H. (1994) 'Structure of peptides associated with class I and class II MHC molecules', *Annu Rev Immunol.*, 12, pp. 181-207, doi: 10.1146/annurev.iy.12.040194.001145.

Etlinger, J.D. and Goldberg, A.L. (1977) 'A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes', *Proc Natl Acad Sci USA*, 74(1) pp. 54-8.

Fan, H., Angelo, G.N., Warren, J.D., Nathan, C.F. and Lin, G. (2014) 'Oxathiazolones Selectively Inhibit the Human Immunoproteasome over the Constitutive Proteasome', *ACS Med Chem Lett.*, 5(4), pp. 405–10.

Farini, A., Sitzia, C., Cassani, B., Cassinelli, L., Rigoni, R., Colleoni, F., Fusco, N., Gatti, S., Bella, P., Villa, C., Napolitano, F., Maiavacca, R., Bosari, S., Villa, A. and Torrente, Y. (2016) 'Therapeutic Potential of Immunoproteasome Inhibition in Duchenne Muscular Dystrophy', *Mol Ther.*, 24(11), pp. 1898-912, doi: 10.1038/mt.2016.162.

Feling, R.H., Buchanan, G.O., Mincer, T.J., Kauffman, C.A., Jensen, P.R. and Fenical, W. (2003) 'Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus Salinospora', *Angew Chem Int Ed Engl*, 42(3), pp. 355-7.

Ferreira, L.G., Dos Santos, R.N., Oliva, G. and Andricopulo, A.D. (2015) 'Molecular docking and structure-based drug design strategies', *Molecules*, 20(7), pp. 13384-421.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P. and Shenkin, P.S. (2004) 'Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy', *J Med Chem*, 47(7), pp. 1739-49.

Gaczynska, M., Osmulski, P.A., Gao, Y., Post, M.J. and Simons, M. (2003) 'Proline- and arginine-rich peptides constitute a novel class of allosteric inhibitors of proteasome activity', *Biochemistry*, 42(29), pp. 8663-70.

Gallastegui, N., Beck, P., Arciniega, M., Huber, R., Hillebrand, S. and Groll, M. (2012) 'Hydroxyureas as noncovalent proteasome inhibitors', *Angew Chem Int Ed Engl.*, 51(1), pp. 247-9, doi: 10.1002/anie.201106010.

Geurink, P.P., van der Linden, W.A., Mirabella, A.C., Gallastegui, N., de Bruin, G., Blom, A.E., Voges, M.J., Mock, E.D., Florea, B.I., van der Marel, G.A., Driessen, C., van der Stelt, M., Groll, M., Overkleeft, H.S. and Kisselev, A.F. (2013) 'Incorporation of non-natural amino acids improves cell permeability and potency of specific inhibitors of proteasome trypsin-like sites', *J Med Chem*, 56(3), pp. 1262-75, doi: 10.1021/jm3016987.

Glickman, M.H. and Ciechanover, A. (2002) 'The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction', *Physiol Rev*, Vols. 82(2), pp. 373-428, doi: 10.1152/physrev.00027.2001.

Goldberg, A. (2003) 'Protein degradation and protection against misfolded or damaged proteins', *Nature*, 426(6968), pp. 895-9.

Goldberg, A.L. (2012) 'Development of proteasome inhibitors as research tools and cancer drugs', *J Cell Biol.*, 199(4), pp. 583-8.

GraphPad Prism version 6.04 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

Gräwert, M.A., Gallastegui, N., Stein, M., Schmidt, B., Kloetzel, P.M., Huber, R. and Groll, M. (2011) 'Elucidation of the α-keto-aldehyde binding mechanism: a lead structure motif for proteasome inhibition', *Angew Chem Int Ed Engl.*, 50(2), pp. 542-4, doi: 10.1002/anie.201005488.

Gräwert, M.A. and Groll, M. (2012) 'Exploiting nature's rich source of proteasome inhibitors as starting points in drug development', *Chem Commun (Camb)*, 48(10), pp. 1364-78, doi: 10.1039/c1cc15273d.

Gregerson, D.A. and Ferrington, D.S. (2012) 'Immunoproteasomes: Structure, Function, and Antigen Presentation', *Prog Mol Biol Transl Sci.*, 109, pp. 75–112, doi: 10.1016/B978-0-12-397863-9.00003-1.

Grimm, S., Ott, C., Hörlacher, M., Weber, D., Höhn, A. and Grune, T. (2012) 'Advanced-glycation-end-product-induced formation of immunoproteasomes: involvement of RAGE and Jak2/STAT1', *Biochem J*, 448(1) pp. 127-39, doi: 10.1042/BJ20120298.

Groll, M., Kim, K.B., Kairies, N., Huber, R. and Crews, C.M. (2000) 'Crystal Structure of Epoxomicin:20S Proteasome Reveals a Molecular Basis for Selectivity of α',β'-Epoxyketone Proteasome Inhibitors', *J. Am. Chem. Soc*, 122(6), pp. 1237–8.

Groll, M., Berkers, C.R., Ploegh, H.L. and Ovaa, H. (2006a) 'Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome', *Structure*, 14(3), pp. 451-6.

Groll, M., Huber, R. and Potts, B.C. (2006b) 'Crystal structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of beta-lactone ring opening and a mechanism for irreversible binding', *J Am Chem Soc.*, 128(15), pp. 5136-41, doi: 10.1021/ja058320b.

Groll, M., Götz, M., Kaiser, M., Weyher, E. and Moroder, L. (2006c) 'TMC-95-based inhibitor design provides evidence for the catalytic versatility of the proteasome', *Chem Biol.*, 13(6), pp. 607-14, doi: 10.1016/j.chembiol.2006.04.005.

Groll, M., Gallastegui, N., Maréchal, X., Le Ravalec, V., Basse, N., Richy, N., Genin, E., Huber, R., Moroder, L., Vidal, J. and Reboud-Ravaux, M. (2010) '20S proteasome inhibition: designing noncovalent linear peptide mimics of the natural product TMC-95A', *Chem Med Chem*, 5(10), pp. 1701-5, doi: 10.1002/cmdc.201000293.

Groll, M., Korotkov, V.S., Huber, E.M., de Meijere, A. and Ludwig, A. (2015) 'A Minimal β-Lactone Fragment for Selective β5c or β5i Proteasome Inhibitors', *Angew Chem Int Ed Engl.*, 54(27), pp. 7810-14, doi: 10.1002/anie.201502931.

Guedes, R.A., Serra, P., Salvador, J.A. and Guedes, R.C. (2016) 'Computational Approaches for the Discovery of Human Proteasome Inhibitors: An Overview', *Molecules*, 21(7), p. 927.

Guex, N. and Peitsch, M.C. (1997) 'SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling', *Electrophoresis*, 18, pp. 2714-23.

Guillaume, B., Chapiro, J., Stroobant, V., Colau, D., Van Holle, B., Parvizi, G., Bousquet-Dubouch, M., Théate, I., Parmentier, N. and van den Eynde, B.J. (2010) 'Two abundant proteasome subtypes that uniquely process some antigens presented by HLA class I molecules', *Proc Natl Acad Sci USA*, 107(43), pp. 18599–604, doi: 10.1073/pnas.1009778107.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E. and Hutchinson, G.R. (2012) 'Avogadro: an advanced semantic chemical editor, visualization, and analysis platform', *J Cheminform*, 4(1), doi: 10.1186/1758-2946-4-17.

Harris, J.L., Alper, P.B., Li, J., Rechsteiner, M. and Backes, B.J. (2001) 'Substrate specificity of the human proteasome', *Chem Biol.*, 8(12), pp. 1131-41.

Harshbarger, W., Miller, C., Diedrich, C. and Sacchettini, J. (2015) 'Crystal structure of the human 20S proteasome in complex with carfilzomib', *Structure*, 23(2), pp. 418-24, doi: 10.1016/j.str.2014.11.017.

Heink, S., Ludwig, D., Kloetzel, P.M. and Krüger, E. (2005) 'IFN-gamma-induced immune adaptation of the proteasome system is an accelerated and transient response', *Proc Natl Acad Sci USA.*, 102(26), pp. 9241-6.

Herndon, T.M., Deisseroth, A., Kaminskas, E., Kane, R.C., Koti, K.M., Rothmann, M.D., Habtemariam, B., Bullock, J., Bray, J.D., Hawes, J., Palmby, T.R., Jee, J., Adams, W., Mahayni, H., Brown, J., Dorantes, A., Sridhara, R., Farrell, A.T. and Pazdur, R. (2013) 'U.S. Food and Drug Administration approval: carfilzomib for the treatment of multiple myeloma', *Clin Cancer Res*, 19(17), pp. 4559-63.

Hess, B., Kutzner, C., Van Der Spoel, D. and Lindahl, E. (2008) 'GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation', *J Chem theory and comp*, *4*(3), pp. 435-447.

Hsu, H.C., Singh, P.K., Fan, H., Wang, R., Sukenick, G., Nathan, C., Lin, G. and Li, H. (2017) 'Structural Basis for the Species-Selective Binding of N,C-Capped Dipeptides to the Mycobacterium tuberculosis Proteasome', *Biochemistry*, 56(1), pp 324–33, doi: 10.1021/acs.biochem.6b01107.

Ho, Y.K., Bargagna-Mohan, P., Wehenkel, M., Mohan, R. and Kim, K.B. (2007) 'LMP2-specific inhibitors: chemical genetic tools for proteasome biology', *Chem Biol.*, 14(4), pp. 419-30.

Huang, X., Luan, B., Wu, J. and Shi, Y. (2016) 'An atomic structure of the human 26S proteasome', *Nat Struct Mol Biol.*, 23(9) pp. 778-85, doi: 10.1038/nsmb.3273.

Huber, E.M., Basler, M., Schwab, R., Heinemeyer, W., Kirk, C.J., Groettrup, M. and Groll, M. (2012) 'Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity', *Cell*, 148(4), pp. 727-38, doi: 10.1016/j.cell.2011.12.030.

Huber, E.M, Heinemeyer, W., de Bruin, G., Overkleeft, H.S. and Groll, M. (2016) 'A humanized yeast proteasome identifies unique binding modes of inhibitors for the immunosubunit β5i', *EMBO J.*, 35(23), pp. 2602-13.

Johnston-Carey, H.K., Pomatto, L.C. and Davies, K.J. (2015) 'The Immunoproteasome in oxidative stress, aging, and disease', *Crit Rev Biochem Mol Biol.*, 51(4), pp. 268-81, doi: 10.3109/10409238.2016.1172554.

Jones, A.K., Woods, A.L., Takeoka, K.T., Shen, X., Wei, J.R., Caughlan, R.E. and Dean, C.R. (2017) 'Determinants of Antibacterial Spectrum and Resistance Potential of the Elongation Factor G Inhibitor Argyrin B in Key Gram-Negative Pathogens', *Antimicrob Agents Chemother*, 61(4), pp. e02400-16. doi: 10.1128/AAC.02400-16.

Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. (1997) 'Development and validation of a genetic algorithm for flexible docking', *J Mol Biol.*, 267(3), pp. 727-48.

Kakkar, T., Pak, Y. and Mayersohn, M. (2000) 'Evaluation of a minimal experimental design for determination of enzyme kinetic parameters and inhibition mechanism', *J Pharmacol Exp Ther.*, 293(3), pp. 861-9.

Kanwar, J., Mohammad, I., Yang, H., Huo, C., Chan, T.H. and Dou, Q.P. (2010) 'Computational modeling of the potential interactions of the proteasome beta5 subunit and catechol-O-methyltransferase-resistant EGCG analogs', *Int J Mol Med.*, 26(2), pp. 209-15.

Kisselev, A.F. and Goldberg, A.L. (2001) 'Proteasome inhibitors: from research tools to drug candidates', *Chem Biol.*, 8(8), pp. 739-58.

Kisselev, A.F., Garcia-Calvo, M., Overkleeft, H.S., Peterson, E., Pennington, M.W., Ploegh, H.L., Thornberry, N.A. and Goldberg, A.L. (2003) 'The caspase-like sites of proteasomes, their substrate specificity, new inhibitors and substrates, and allosteric interactions with the trypsin-like sites', *J Biol Chem.*, 278(38), pp. 35869-77.

Kisselev, A.F. and Goldberg, A.L. (2005) 'Monitoring activity and inhibition of 26S proteasomes with fluorogenic peptide substrates', *Methods Enzymol.*, 398, pp. 364-78.

Kisselev, A.F., Callard, A. and Goldberg, A.L. (2006) 'Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate', *J Biol Chem.*, 281(13), pp. 8582-90.

Kisselev, A.F., van der Linden, W.A., and Overkleeft, H.S. (2012) 'Proteasome Inhibitors: An Expanding Army Attacking a Unique target', *Chem Biol*, 19(1), pp. 99–115, doi:10.1016/j.chembiol.2012.01.003.

Koroleva, O.N., Pham, T.H., Bouvier, D., Dufau, L., Qin, L., Reboud-Ravaux, M., Ivanov, A.A., Zhuze, A.L., Gromova, E.S. and Bouvier-Durand, M. (2015) 'Bisbenzimidazole derivatives as potent inhibitors of the trypsin-like sites of the immunoproteasome core particle', *Biochimie*, 108, pp. 94-100. doi: 10.1016/j.biochi.2014.11.002.

Kramer, B., Rarey, M. and Lengauer, T. (1999) 'Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking', *Proteins*, 37(2), pp. 228-41.

Krunic, A., Vallat, A., Mo, S., Lantvit, D.D., Swanson, S.M. and Orjala, J. (2010) 'Scytonemides A and B, cyclic peptides with 20S proteasome inhibitory activity from the cultured cyanobacterium Scytonema hofmanii', *J Nat Prod*, 73(11), pp. 1927-32, doi: 10.1021/np100600z.

Kuhn ,D.J., Hunsucker, S.A., Chen, Q., Voorhees, P.M., Orlowski, M. and Orlowski, R.Z.(2009) 'Targeted inhibition of the immunoproteasome is a potent strategy against models of multiple myeloma that overcomes resistance to conventional drugs and nonspecific proteasome inhibitors', *Blood*, 113(19), pp. 4667-76, doi: 10.1182/blood-2008-07-171637.

Lee, D. and Goldberg, A.L.(1998) 'Proteasome inhibitors: valuable new tools for cell biologists', *Trends Cell Biol.*, 8(10), pp. 397-403.

Li, H., O'Donoghue, A.J., van der Linden, W.A., Xie, S.C., Yoo, E., Foe, I.T., Tilley, L., Craik, C.S., da Fonseca, P.C. and Bogyo, M. (2016) 'Structure- and function-based design of Plasmodium-selective proteasome inhibitors', *Nature*, 530(7589), pp. 233-6, doi: 10.1038/nature16936.

Li, X., Wood, T.E., Sprangers, R., Jansen, G., Franke, N.E., Mao, X., Wang, X., Zhang, Y., Verbrugge, S.E., Adomat, H., Li, Z.H., Trudel, S., Chen, C., Religa, T.L., Jamal, N., Messner, H., Cloos, J., Rose, D.R., Navon, A., Guns, E., Batey, R.A., Kay, L.E. and Schimmer, A.D. (2010) 'Effect of noncompetitive proteasome inhibition on bortezomib resistance', *J Natl Cancer Inst.*, 102(14), pp. 1069-82, doi: 10.1093/jnci/djq198.

Lightcap, E.S., McCormack, T.A., Pien, C.S., Chau, V., Adams, J. and Elliott, P.J. (2000) 'Proteasome Inhibition Measurements: Clinical Application', *Clin Chem*, 46(5), pp. 673-83.

Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) 'Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings', *Adv Drug Deliv Rev*, 46(1-3), pp. 3-26.

Loizidou, E.Z. and Zeinalipour-Yazdi, C.D. (2014) 'Computational inhibition studies of the human proteasome by argyrin-based analogues with subunit specificity', *Chem Biol Drug Des*, 84(1), pp. 99-107.

Manasanch, E.E. and Orlowski, R.Z. (2017) 'Proteasome inhibitors in cancer therapy', *Nat Rev Clin Oncol.*, 14(7), pp. 417-33, doi: 10.1038/nrclinonc.2016.206.

Mason, J.S., Bortolato, A., Weiss, D.R., Deflorian, F., Tehan, B. and Marshall, F.H. (2013) 'High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks', *In silico pharma*, *1*(1), p.23.

McGann, M. (2011) 'FRED pose prediction and virtual screening accuracy', *J Chem Inf Model*, 51(3), pp. 578-96, doi: 10.1021/ci100436p.

McGann, M. (2012) 'FRED and HYBRID docking performance on standardized datasets', *J. Comput. Aided Mol. Des*, 26(8), pp. 897-906, doi: 10.1007/s10822-012-9584-8.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P. and Lopez, R. (2013) 'Analysis tool web services from the EMBL-EBI', *Nucleic Acids Res*, 41(W1), pp. W597-W600, doi: 10.1093/nar/gkt376.

Micale, N., Scarbaci, K., Troiano, V., Ettari, R., Grasso, S. and Zappalà, M. (2014) 'Peptide-Based Proteasome Inhibitors in Anticancer Drug Design', *Med Res Rev*, 34(5), pp. 1001-69.

Miller, Z., Ao, L., Kim, K.B. and Lee, W. (2013) 'Inhibitors of the immunoproteasome: current status and future directions', *Curr Pharm Des.*, 19(22), pp. 4140-51.

Minitab 17.1.0 Statistical Software (2010). [Computer software]. State College PA: Minitab, Inc. (www.minitab.com).

Mishto, M., Bellavista, E., Santoro, A., Stolzing, A., Ligorio, C., Nacmias, B., Spazzafumo, L., Chiappelli, M., Licastro, F., Sorbi, S., Pession, A., Ohm, T., Grune, T. and Franceschi, C. (2006) 'Immunoproteasome and LMP2 polymorphism in aged and Alzheimer's disease brains', *Neurobiol Aging.*, 27(1), pp. 54-66.

Mohan, C., Long, K. and Mutneja, M. (2013) 'An introduction to inhibitors and their biological applications', *EMD Millipore Corp*, 1st edition.

Moreau, P., Masszi, T., Grzasko, N., Bahlis, N.J., Hansson, M., Pour, L., Sandhu, I., Ganly, P., Baker, B.W., Jackson, S.R., Stoppa, A.M., Simpson, D.R., Gimsing, P., Palumbo, A., Garderet, L., Cavo, M., Kumar, S., Touzeau, C., Buadi, F.K., Laubach, J.P., Berg, D.T., Lin, J. and Di Bacco, A., Hui, A.M., Van De Velde, H., Richardson, P.G., Kupperman, E., Berger, A., Dick, L., Williamson, M., Chauhan, D., Anderson, K., Esseltine, D.L., Niculescu, L., Taraskiewicz, S., Stecklair, J., Lonial, S., Rajkumar, V., Miquel, J.S., Einsele, H. and Orlowski, R. (2016) 'Oral Ixazomib, Lenalidomide, and Dexamethasone for Multiple Myeloma', *N Engl J Med*, 374(17), pp. 1621-34, doi: 10.1056/NEJMoa1516282.

Morris, G., Huey, R., Lindstrom, W., Sanner, M., Belew, R., Goodsell, D. and Olson, A. (2009) 'Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity', *J. Computational Chemistry*, 16, pp. 2785-91.

Motulsky, H. and Christopolous, A. (2003) 'Fitting models to biological data using linear and nonlinear regression - A practical guide to curve fitting', *San Diego : GraphPad Software Inc*, Vol. 4.0.

Muchamuel, T., Basler, M., Aujay, M.A., Suzuki, E., Kalim, K.W., Lauer, C., Sylvain, C., Ring, E.R., Shields, J., Jiang, J. and Shwonek, P., Parlati, F., Demo, S.D., Bennett, M.K., Kirk, C.J., Groettrup, M., (2009) 'A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis', *Nat Med*, 15(7), pp.781-787.

Mundy, G., Gutierrez, G., Garrett, R., Gallwitz, W., Rossini, G., Christiansen, C. and Langenberg, A. (2007) 'Proteasome inhibitors stimulate both bone formation and hair growth by similar mechanisms', *Ann N Y Acad Sci*, 1117, pp. 298-301.

Nickeleit, I., Zender, S., Sasse, F., Geffers, R., Brandes, G., Sörensen, I., Steinmetz, H., Kubicka, S., Carlomagno, T., Menche, D., Gütgemann, I., Buer, J., Gossler, A., Manns, M.P., Kalesse, M., Frank, R. and Malek, N.P. (2008) 'Argyrin A reveals a critical role for the tumor suppressor protein p27(kip1) in mediating antitumor activities in response to proteasome inhibition', *Cancer Cell*, 14(1), pp. 23-35.

Nitta, T., Murata, S., Sasaki, K., Fujii, H., Ripen, A.M., Ishimaru, N., Koyasu, S., Tanaka, K. and Takahama, Y. (2010) 'Thymoproteasome Shapes Immunocompetent Repertoire of CD8+ T Cells', *Immunity*, 32(1), pp. 29-40, doi: 10.1016/j.immuni.2009.10.009.

Nyfeler, B., Hoepfner, D., Palestrant, D., Kirby, C.A., Whitehead, L., Yu, R., Deng, G., Caughlan, R.E., Woods, A.L., Jones, A.K., Barnes, S.W., Walker, J.R., Gaulis, S., Hauy, E., Brachmann, S.M., Krastel, P., Studer, C., Riedl, R., Estoppey, D., Aust, T., Movva, N.R., Wang, Z., Salcius, M., Michaud, G.A., McAllister, G., Murphy, L.O., Tallarico, J.A., Wilson, C.J. and Dean, C.R. (2012) 'Identification of elongation factor G as the conserved cellular target of argyrin B', *PLoS One*, 7(9), pp. e42657, doi: 10.1371/journal.pone.0042657.

Palatinus, L., Brázda, P., Boullay, P., Perez, O., Klementová, M., Petit, S., Eigner, V., Zaarour, M. and Mintova, S. (2017) 'Hydrogen positions in single nanocrystals revealed by electron diffraction', *Science*, 355(6321), pp. 166-9, doi: 10.1126/science.aak9652.

Palombella, V.J., Rando, O.J., Goldberg, A.L. and Maniatis, T. (1994) 'The ubiquitinproteasome pathway is required for processing the NF-κB1 precursor protein and the activation of NF-κB', *Cell*, 78(5), pp. 773-85.

Pandey, U.B., Nie, Z., Batlevi, Y., McCray, B.A., Ritson, G.P., Nedelsky, N.B., Schwartz, S.L., Di Prospero, N.A., Knight, M.A., Schuldiner, O., Padmanabhan, R., Hild, M., Berry, D.L., Garza, D., Hubbert, C.C., Yao, T.P., Baehrecke, E.H. and Taylor, J.P. (2007) 'HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS', *Nature*, 447(7146), pp. 859-63.

Parlati, F., Lee, S.J., Aujay, M., Suzuki, E., Levitsky, K., Lorens, J.B., Micklem, D.R., Ruurs, P., Sylvain, C., Lu, Y., Shenk, K.D. and Bennett, M.K. (2009) 'Carfilzomib can induce tumor cell death through selective inhibition of the chymotrypsin-like activity of the proteasome', *Blood*, 114(16), pp. 3439-47.

Pellom Jr, S.T. and Shanker, A. (2012) 'Development of Proteasome Inhibitors as Therapeutic Drugs', *J Clin Cell Immunol.*, 15(S5), p. 5.

Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D. and Goldberg, A.L. (1994) 'Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules', *Cell*, 78(5), pp. 761-71.

Sasse, F., Steinmetz, H., Schupp, T., Petersen, F., Memmert, K., Hofmann, H., Heusser, C., Brinkmann, V., von Matt, P., Höfle, G. and Reichenbach, H. (2002), 'Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties', *J Antibiot (Tokyo)*, 55(6), pp. 543-51.

Schrader, J., Henneberg, F., Mata, R.A., Tittmann, K., Schneider, T.R., Stark, H., Bourenkov, G. and Chari, A. (2016) 'The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design', *Science*, 353(6299), pp. 594-8, doi: 10.1126/science.aaf8993.

Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A. and Farid, R. (2006) 'Novel procedure for modeling ligand/receptor induced fit effects', *J med chem*, *49*(2), pp.534-553.

Shi, Y.Y., Small, G.W. and Orlowski, R.Z. (2006) 'Proteasome inhibitors induce a p38 mitogen-activated protein kinase (MAPK)-dependent anti-apoptotic program involving MAPK phosphatase-1 and Akt in models of breast cancer', *Breast Cancer Res Treat.*, 100(1), pp. 33-47.

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J. and Sherman, W. (2010) 'Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field', *J chem theory comp*, *6*(5), pp.1509-1519.

Siegel, D.S., Martin, T., Wang, M., Vij, R., Jakubowiak, A.J., Lonial, S., Trudel, S., Kukreti, V., Bahlis, N., Alsina, M., Chanan-Khan, A., Buadi, F., Reu, F.J., Somlo, G., Zonder, J., Song, K., Stewart, A.K., Stadtmauer, E., Kunkel, L., Wear, S., Wong, A.F., Orlowski, R.Z. and Jagannath, S. (2012) 'A phase 2 study of single-agent carfilzomib (PX-171-003-A1) in patients with relapsed and refractory multiple myeloma', *Blood*, 120(14), pp. 2817-25.

Singh, A.V., Bandi, M., Aujay, M.A., Kirk, C.J., Hark, D.E., Raje, N., Chauhan, D. and Anderson, K.C. (2010) 'PR-924, a selective inhibitor of the immunoproteasome subunit LMP-7, blocks multiple myeloma cell growth both in vitro and in vivo', *Br J Haematol.*, 152(2), pp. 155-63, doi: 10.1111/j.1365-2141.2010.08491.

Sosič, I., Gobec, M., Brus, B., Knez, D., Živec, M., Konc, J., Lešnik, S., Ogrizek, M., Obreza, A., Žigon, D., Janežič, D., Mlinarič-Raščan, I. and Gobec, S. (2016) 'Nonpeptidic Selective Inhibitors of the Chymotrypsin-Like (β5 i) Subunit of the Immunoproteasome', *Angew Chem Int Ed Engl.*, 55(19), pp. 5745-8, doi: 10.1002/anie.201600190.

Sousa, S.F., Ribeiro, A.J., Coimbra, J.T., Neves, R.P., Martins, S.A., Moorthy, N.S., Fernandes, P.A. and Ramos, M.J. (2013) 'Protein-ligand docking in the new millennium - a retrospective of 10 years in the field', *Curr Med Chem.*, 20(18), pp. 2296-314.

Spence, J.T., Cotton, J.W., Underwood, B.J. and Duncan, C.P. (1976) 'Elementary Statistics 3rd edition', *Prentice-Hall Englewood-Cliffs*, 3rd edition.

Spiess, A.N. and Neumeyer, N. (2010) 'An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach', *BMC Pharmacol.*, 10(1), p. 6, doi: 10.1186/1471-2210-10-6.

Spitzer, R. and Jain, A.N. (2012) 'Surflex-Dock: Docking benchmarks and real-world application', *J Comput Aided Mol Des*, 26(6), pp. 687-99, doi: 10.1007/s10822-011-9533-y.

Stadelmann, B., Aeschbacher, D., Huber, C., Spiliotis, M., Müller, J. and Hemphill, A. (2014) 'Profound Activity of the Anti-cancer Drug Bortezomib against Echinococcus multilocularis Metacestodes Identifies the Proteasome as a Novel Drug Target for Cestodes', *PLoS Negl Trop Dis*, 8(12), e3352.

Stank, A., Kokh, D.B., Horn, M., Sizikova, E., Neil, R., Panaecka, J., Richter, S. and Wade, R. (2017) 'TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets', *Nucl Acids Res*, 45(W1), pp. W325-W330, doi: 10.1093/nar/gkx277.

Stauch, B., Simon, B., Basile, T., Schneider, G., Malek, N.P., Kalesse, M. and Carlomagno, T. (2010) 'Elucidation of the structure and intermolecular interactions of a reversible cyclic-peptide inhibitor of the proteasome by NMR spectroscopy and molecular modeling', *Angew Chem Int Ed Engl*, 49(23), pp. 3934-8, doi: 10.1002/anie.201000140.

Strelow, J., Dewe, W., Iversen, P.W., Brooks, H.B., Radding, J.A., McGee, J. and Weidner, J. (2012) 'Mechanism of Action Assays for Enzymes', *Advancing Translational Sciences*.

The PyMOL Molecular graphics System, Version 1.7.4.5, Schrödinger, LLC.

Tiptona, K.F., Armstrong, R.N., Bakker, B.M., Bairoch, A., Cornish-Bowden, A., Halling, P.J., Hofmeyr, J.H., Leyh, T.S., Kettner, C., Raushel, F.M., Rohwer, J., Schomburg, D. and Steinbeck, C. (2014) 'Standards for Reporting Enzyme Data: The STRENDA Consortium: What it aims to do and why it should be helpful', *Perspectives in Science*, 1(6), pp. 131-7.

UniProt Consortium. (2015) 'UniProt: a hub for protein information', *Nucl. Acids Res.*, 43(D), pp. D204-D212, doi: 10.1093/nar/gku989.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. and Taylor, R.D. (2003) 'Improved protein-ligand docking using GOLD', *Proteins*, 52(4), pp. 609-23.

Vollbrecht, L., Steinmetz, H., Hofle, G., Oberer, L., Rihs, G., Bovermann, G. and von Matt, P. (2002) 'Argyrins, immunosuppressive cyclic peptides from myxobacteria. II. Structure elucidation and stereochemistry', *J Antibiot (Tokyo)*, 55(8), pp. 715-21.

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M.K., Greenwood, J. and Romero, D.L. (2015) 'Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field', *J Am Chem Soc*, *137*(7), pp.2695-2703.

Wang, R., Lai, L. and Wang, S.. (2002) 'Further development and validation of empirical scoring functions for structure-based binding affinity prediction', *J Comput Aided Mol Des*, 16(1), pp. 11-26.

Wildeboer, D., Jeganathan, F., Price, R.G. and Abuknesha, R.A. (2008) 'Characterization of bacterial proteases with a panel of fluorescent peptide substrates', *Anal. Biochem.*, 384(2), pp. 321-8, doi: 10.1016/j.ab.2008.

Yang, H.,Landis-Piwowar, K.R., Lu, D., Yuan, P., Li, L., Reddy, G.P., Yuan, X. and Dou, Q.P. (2008) 'Pristimerin induces apoptosis by targeting the proteasome in prostate cancer cells', *J Cell Biochem.*, 103(1), pp. 234-44.

Zhu, K., Dunner, K.Jr. and McConkey, D.J. (2010) 'Proteasome inhibitors activate autophagy as a cytoprotective response in human prostate cancer cells', *Oncogene*, 29(3), pp. 451-62, doi: 10.1038/onc.2009.343.

Zoete, V., Cuendet, M.A., Grosdidier, A. and Michielin, O. (2011) 'SwissParam: a fast force field generation tool for small organic molecules', *J Comp Chem*, *32*(11), pp.2359-2368.

Chapter 5. Appendix

List of Supplementary Figures

List of Supplementary Tables

Supplementary table 1- Stock, working and final concentrations of reagents used.

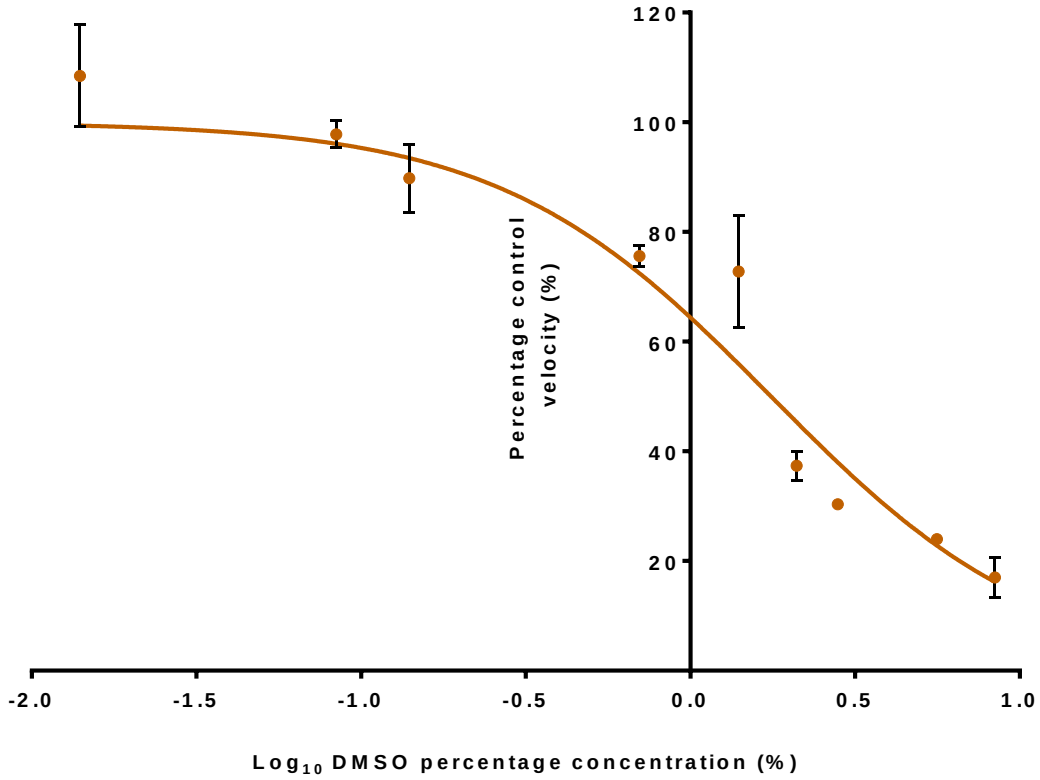
Reagent	Stock concentration	Working concentration	Final concentration
Argyrin B	10 mg/ml in DMSO	As required, in proteasome buffer	As required, in proteasome buffer
β1c substrate	37.5 mM in DMSO	375-500 µM, in proteasome buffer	As required, in proteasome buffer
β1i substrate	37.5 mM in DMSO	375-500 µM, in proteasome buffer	As required, in proteasome buffer
β5c / β5i substrate	37.5 mM in DMSO	375-500 µM, in proteasome buffer	As required, in proteasome buffer
Constitutive proteasome	1 mg/ml	10 µg/ml	0.1 µg/well
Epoxomycin	500 µM in DMSO	As required, in proteasome buffer	As required, in proteasome buffer
Immunoproteasome	1 mg/ml	10 µg/ml	0.1 µg/well
Vinyl Sulfone	500 µM in DMSO	As required, in proteasome buffer	As required, in proteasome buffer

Supplementary Table 2 – Lamarckian 4.2 genetic algorithm parameters used in AutoDock simulations.

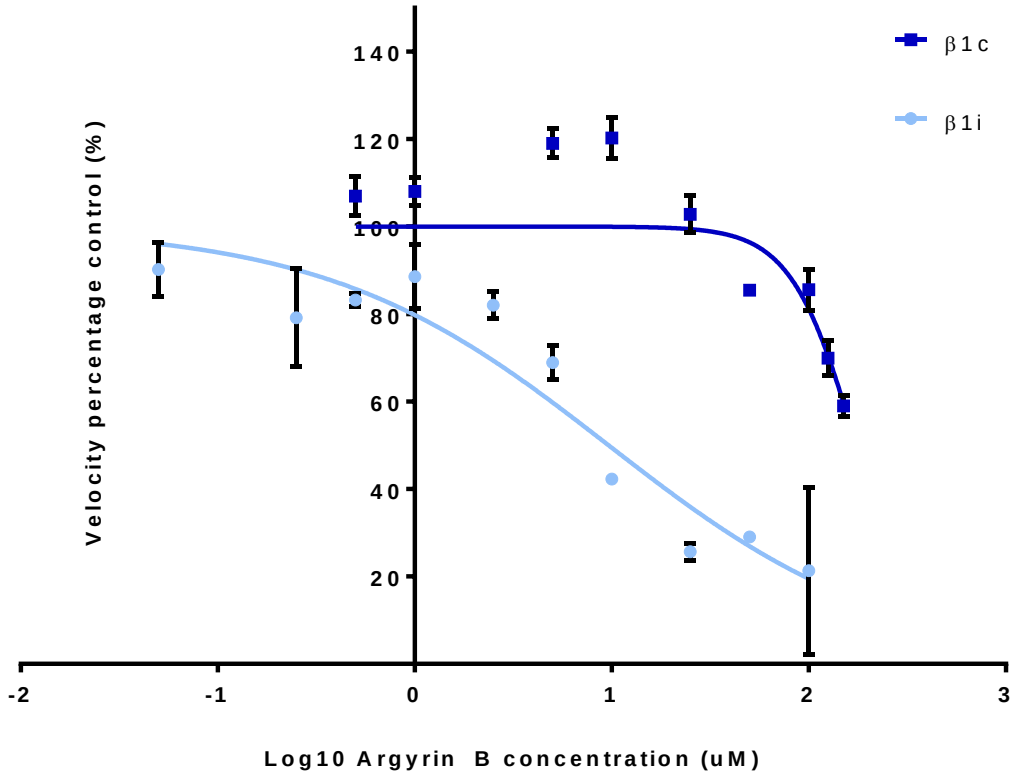
Genetic Algorithm Parameters	
Number of runs per test	50
Population size	150
Maximum number of evals	2500000
Maximum number of generations	27000
Maximum number of top individuals that automatically survive	1
Rate of gene mutation	0.02
Rate of crossover	0.8
Genetic algorithm crossover mode	Two pt

Supplementary table 3 - Chain and coordinates of grid box centre for each active site cut – Structural data from PDB:4R30 for human CP, PDB:3UNH for human IP and PDB:1JD2 for yeast CP. 'h' represents humanised, 'y' represent yeast.

Active Site	Chain	Threonine 1 coordinate positioning of grid box		
		x	y	z
β1c	H	-44.697	77.350	-80.590
β2c	I	-59.837	60.333	-97.139
β5c	Z	-42.749	25.372	-104.397
hβ1i	N	23.238	-78.082	-11.152
hβ2i	H	3.687	-59.799	-18.031
hβ5i	K	52.961	-24.734	-1.942
yβ2c	o	4.857	-159.860	43.634



Supplementary figure 1 – Effect of solvent on rate of IP reaction – Increasing DMSO solvent concentration on a logarithmic scale, against the percentage of control IP reaction rate. β1i purified assay with Ac-PAL-AMC substrate. Nonlinear regression analysis fit with normalised response and variable slope. Data presented as average with standard error bars.



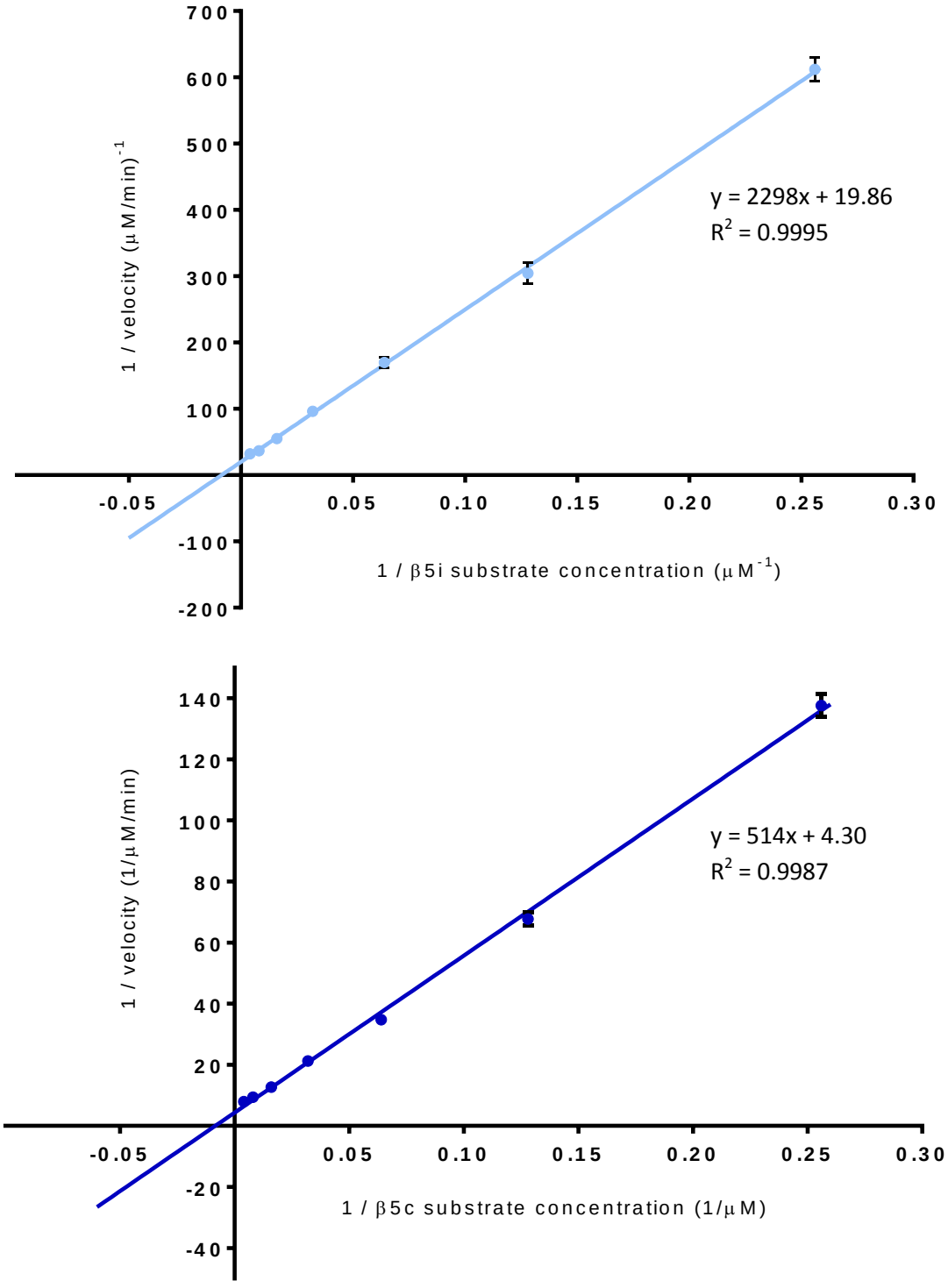
Supplementary figure 2 – β1c and β1i IC_{50} estimation – Tested at [CP] and [IP] = 0.1 µg/well and [S] = 50 µM. Analysed using SNLR variable slope. β1i IC_{50} 9.6 µM +/- 1.5 and β1c IC_{50} = 173.9 µM +/-1.1.

Supplementary table 4- β1c K_{ic} values determined from Dixon plot – Calculated from intersecting lines of different substrate concentrations.

β1i [S] (µM)	10.9	21.9	43.8	87.5	175
10.9		2.5	3.67	3.68	3.68
21.9	2.50		5.85	5.18	5.00
43.8	3.67	5.85		3.78	3.76
87.5	3.68	5.18	3.78		3.72
175	3.68	5.00	3.76	3.72	

Supplementary table 5 – β1i K_{iu} values determined from Cornish-Bowden plot – Calculated from intersecting lines of different substrate concentrations.

β1i [S] (µM)	10.9	21.9	43.8	87.5	175
10.9		30.58	13.24	13.07	10.97
21.9	30.58		7.96	7.43	9.06
43.8	13.24	7.96		12.93	10.54
87.5	13.07	7.43	12.93		9.77
175	10.97	9.06	10.54	9.77	



Supplementary figure 3 - β5c and β5i Lineweaver-Burk plots to estimate K_m and V_{max} – Linear transformations with line equations solved to determine axis intercepts. β5c: V_{max} = 0.2320 μM/min, K_m = 119.3 μM. Excluding 3.9 μM data point K_m = 89.9 μM. β5i: V_{max} = 0.05035 μM/min, K_m = 115.714 μM. Excluding 3.9 μM data point K_m = 100.1 μM.

Supplementary table 6 - Unpaired t-test p values calculated from IC_{50} mean from independent 3 replicates - Bold denotes statistical significance between active sites, at p<0.05.

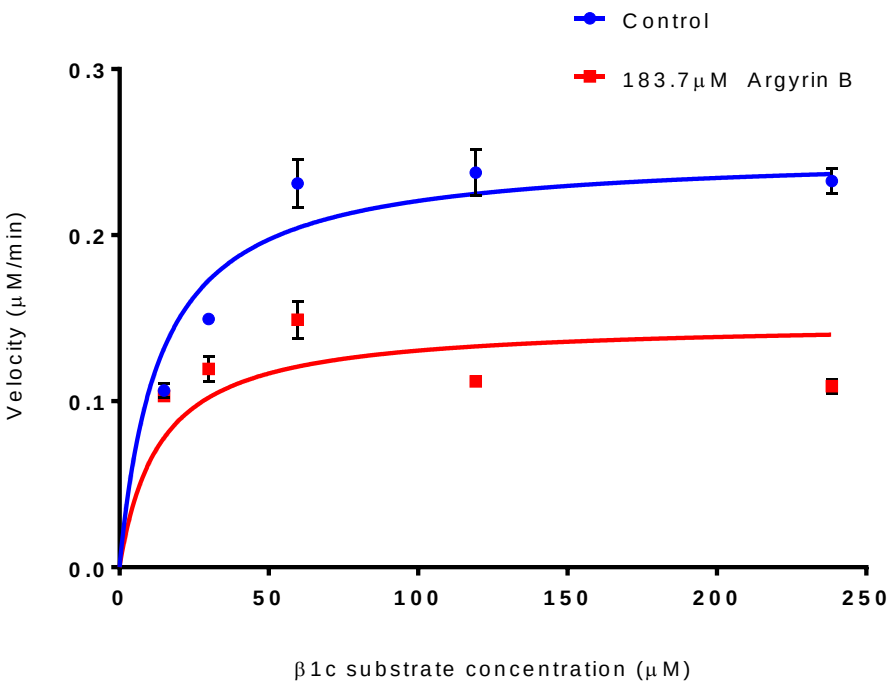
	β1c	β1i	β5c	β5i
β1c		**0.0002**	**0.0002**	**0.0001**
β1i	**0.0002**		0.8586	0.0744
β5c	**0.0002**	0.8586		0.1015
β5i	**0.0001**	0.0744	0.1015	

Supplementary table 7 – Best-fit model analysis results from Akaike's information criteria.

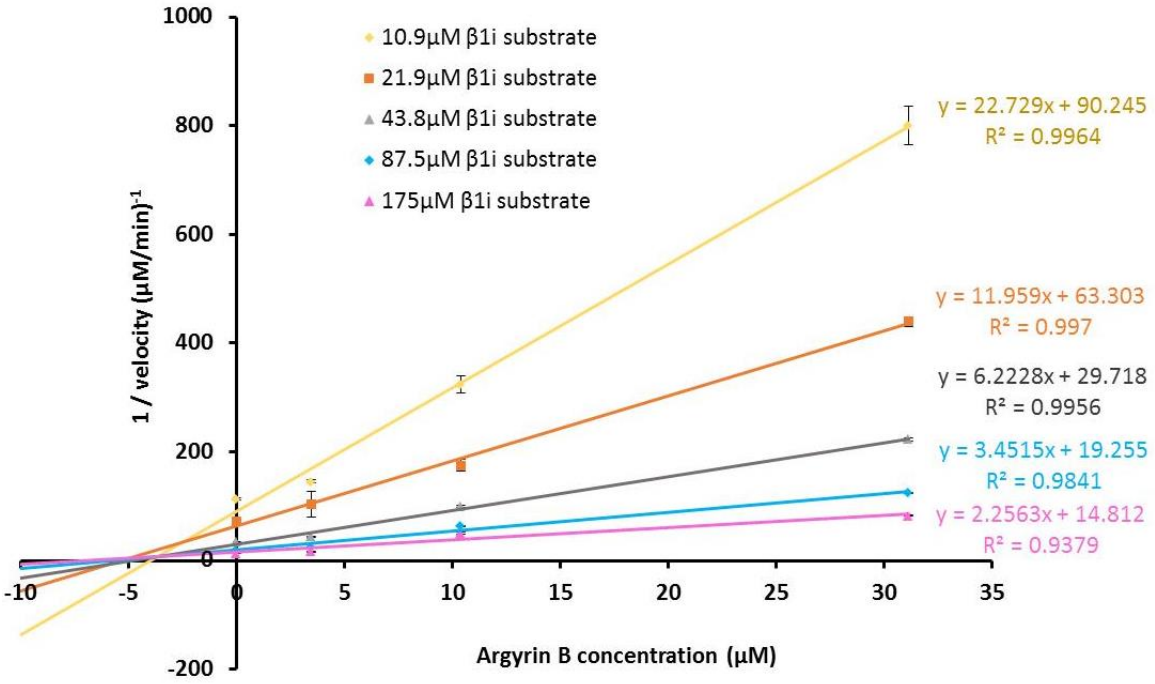
Active Site	Probability (%)			AICc Difference
	Competitive	Mixed	Non-competitive	
β1i	0.01	99.99		18.10
	<0.01		>99.99	20.38
		24.2	75.8	-2.28
β5c	<0.01	>99.99		45.31
	<0.01		>99.99	44.57
		59.19	40.81	0.74
β5i	<0.01	>99.99		31.44
	<0.01		>99.99	33.76
		23.86	76.14	-2.32

Supplementary table 8 – Best-fit model analysis from extra sum of squares F test at p<0.05.

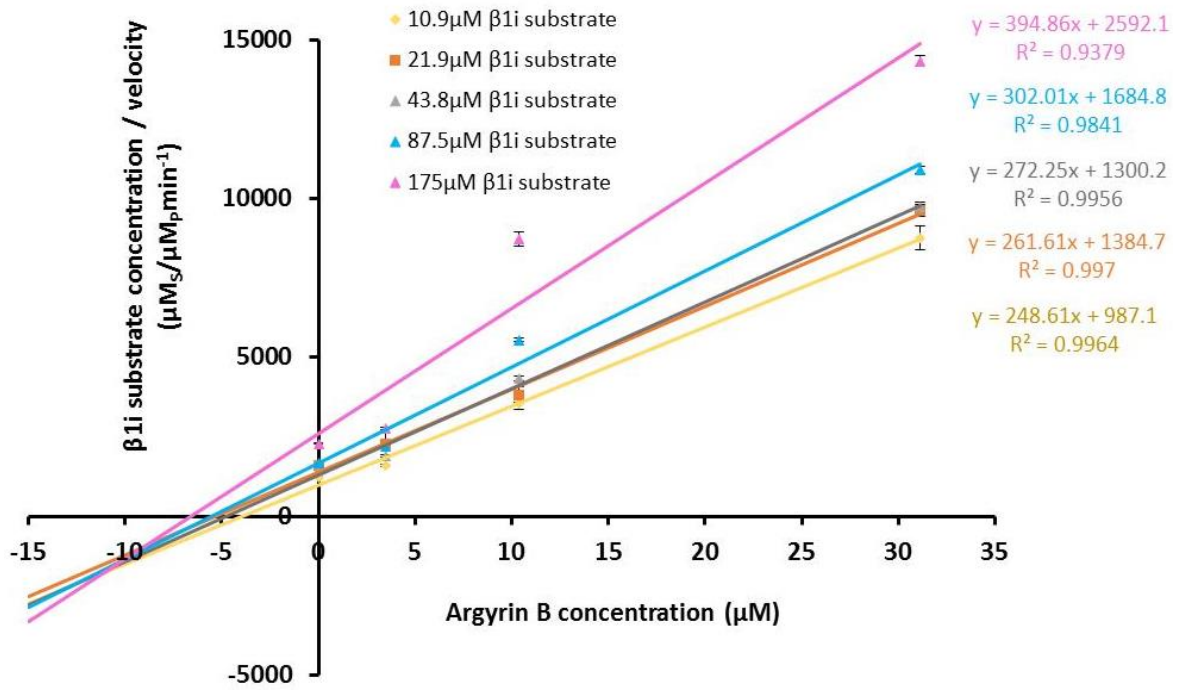
Active Site	Mixed	Non-competitive	p-value	F ratio
β1i	reject	accept	0.7609	0.09352
β5c	reject	accept	0.089	2.996
β5i	reject	accept	0.8099	0.05841



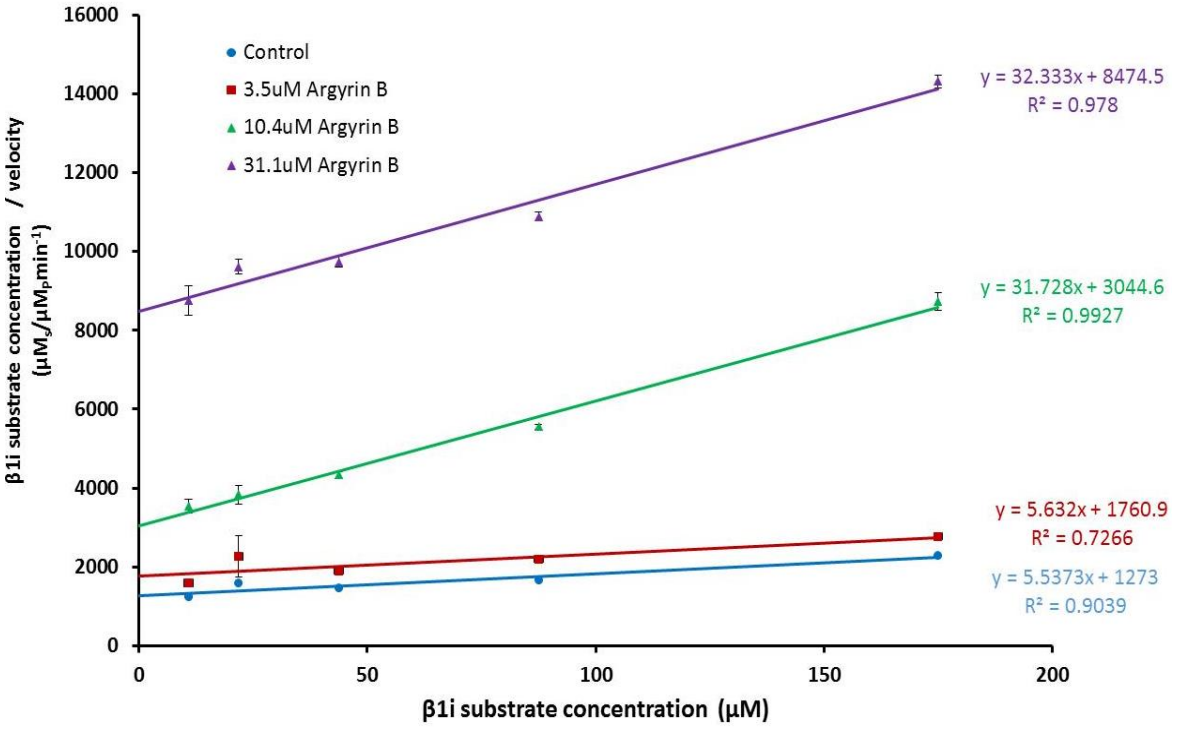
Supplementary figure 4 – Argyrin B inhibition at β1c - Tested at 183.7 µM argyrin B with DMSO correction equal to 1.54% DMSO. [CP] = 0.1 µg/well and SNLR analysis fit with non-competitive model. Estimated K_i of 265 µM from single inhibitor concentration.



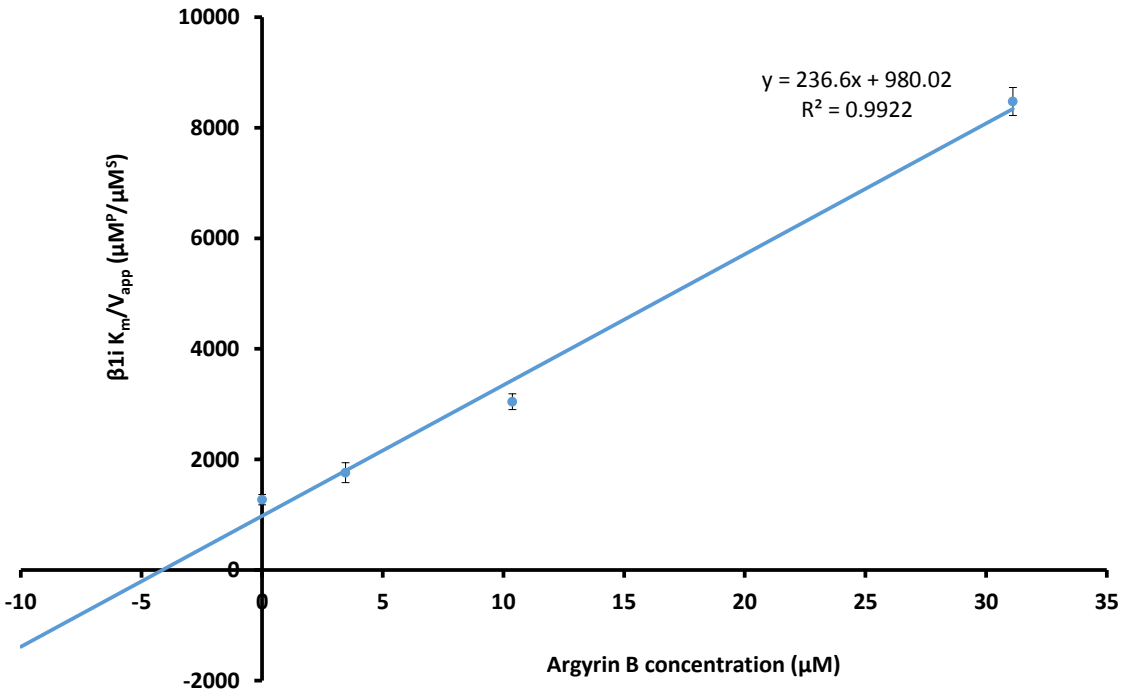
Supplementary figure 5 – Dixon plot with range of argyrin B and substrate concentrations at β1i – Intersecting lines estimate K_{ic} = 4.08 µM.



Supplementary figure 6 - Cornish Bowden plot with range of argyrin B and substrate concentrations at β1i – Intersecting lines estimate K_{iu} = 10.55 µM.



Supplementary figure 7 - Hanes-Woolf plot of [S]/v against [S] for argyrin B at β1i - Y intercept values represent K_m/V_{app} for each inhibitor concentration, used to estimate K_i.



Supplementary figure 8 – K_i estimate from K_m/V_{app} against [I] for argyrin B at β1i – K_i value from x-intercept = 4.14 μM.

LMP-2 [Homo sapiens]

Sequence ID: gb|AAC50154.1| **Length:** 209 **Number of Matches:** 2

▸ See 1 more title(s)

Range 1: 11 to 209 GenPept Graphics ▼ Next Match ▲ Previous M

Score	Expect	Method	Identities	Positives	Gaps
372 bits(955)	5e-125	Compositional matrix adjust.	179/199(90%)	193/199(96%)	1/199(0%)

```
Query  189  TTIMAVEFDGGVVVGSDSRVSAGTAVVNRVFDKLSPLHQRIFCALSGSAADAQAIADMAA  248
            TTIMAVEFDGGVV+GSDSRVSAG AVVNRVFDKLSPLH+RI+CALSGSAADAQA+ADMAA
Sbjct  11   TTIMAVEFDGGVVMGSDSRVSAGEAVVNRVFDKLSPLHERIYCALSGSAADAQAVADMAA  70

Query  249  YQLELHGLELE-PPLVLAAANVVKNISYKYREDLLAHLIVAGWDQREGGQVYGTMGGMLI  307
            YQLELHG+ELE PPLVLAAANVV+NISYKYREDL AHL+VAGWDQREGGQVYGT+GGML
Sbjct  71   YQLELHGIELEEPPLVLAAANVVRNISYKYREDLSAHLMVAGWDQREGGQVYGTLGGMLT  130

Query  308  RQPFTIGGSGSSYIYGYVDAAYKPGMTPEECRRFTTNAITLAMNRDGSSGGVIYLVTITA  367
            RQPF IGGSGS++IYGYVDAAYKPGM+PEECRRFTT+AI LAM+RDGSSGGVIYLVTITA
Sbjct  131  RQPFAIGGSGSTFIYGYVDAAYKPGMSPEECRRFTTDAIALAMSRDGSSGGVIYLVTITA  190

Query  368  AGVDHRVILGDELPKFYDE  386
            AGVDHRVILG+ELPKFYDE
Sbjct  191  AGVDHRVILGNELPKFYDE  209
```

Supplementary Figure 9 - BLASTp alignment from murine β1i active site cut (query), to human sequence β1i (subject) – Murine IP sequence from PDB:3UNH, human β1i sequence from UniProt:P28065.

proteasome subunit beta type-10 precursor [Homo sapiens]

Sequence ID: ref|NP_002792.1| **Length:** 273 **Number of Matches:** 1

▸ See 9 more title(s)

Range 1: 40 to 239 GenPept Graphics ▼ Next Match ▲ Previous

Score	Expect	Method	Identities	Positives	Gaps
358 bits(920)	1e-118	Compositional matrix adjust.	177/200(89%)	185/200(92%)	4/200(2%)

```
Query  18   TTIAGLVFRDGVILGADTRATNDSVVADKSCEKIHFIAPKIYCCGAGVAADTEMTTRMAA  77
            TTIAGLVF+DGVILGADTRATNDSVVADKSCEKIHFIAPKIYCCGAGVAAD EMTTRM A
Sbjct  40   TTIAGLVFQDGVILGADTRATNDSVVADKSCEKIHFIAPKIYCCGAGVAADAEMTTRMVA  99

Query  78   SKMELHA----REPRVATVTRILRQTLFRYQGHVGASLVVGGVDLNGPQLYEVHPHGSYS  133
            SKMELHA    REPRVATVTRILRQTLFRYQGHVGASL+VGGVDL GPQLY VHPHGSYS
Sbjct  100  SKMELHALSTGREPRVATVTRILRQTLFRYQGHVGASLIVGGVDLTGPQLYGVHPHGSYS  159

Query  134  RLPFTALGSGQGAAVALLEDRFQPNMTLEAAQELLVEAITAGILSDLGSGGNVDACVITA  193
            RLPFTALGSGQ AA+A+LEDRFQPNMTLEAAQ LLVEA+TAGIL DLGSGGNVDACVIT
Sbjct  160  RLPFTALGSGQDAALAVLEDRFQPNMTLEAAQGLLVEAVTAGILGDLGSGGNVDACVITK  219

Query  194  GGAKLQRALSTPTEPVQRAG  213
             GAKL R LS+PTEPV+R+G
Sbjct  220  TGAKLLRTLSSPTEPVKRSG  239
```

Supplementary Figure 10 - BLASTp alignment from murine β2i active site cut (query), to human sequence β2i (subject) – Murine IP sequence from PDB:3UNH, human β1i sequence from UniProt:P40306.

proteasome subunit beta type-8 isoform E1 precursor [Homo sapiens]

Sequence ID: ref|NP_004150.1| **Length:** 272 **Number of Matches:** 1

▸ See 7 more title(s)

Range 1: 69 to 269 GenPept Graphics ▼ Next Match ▲ Previous

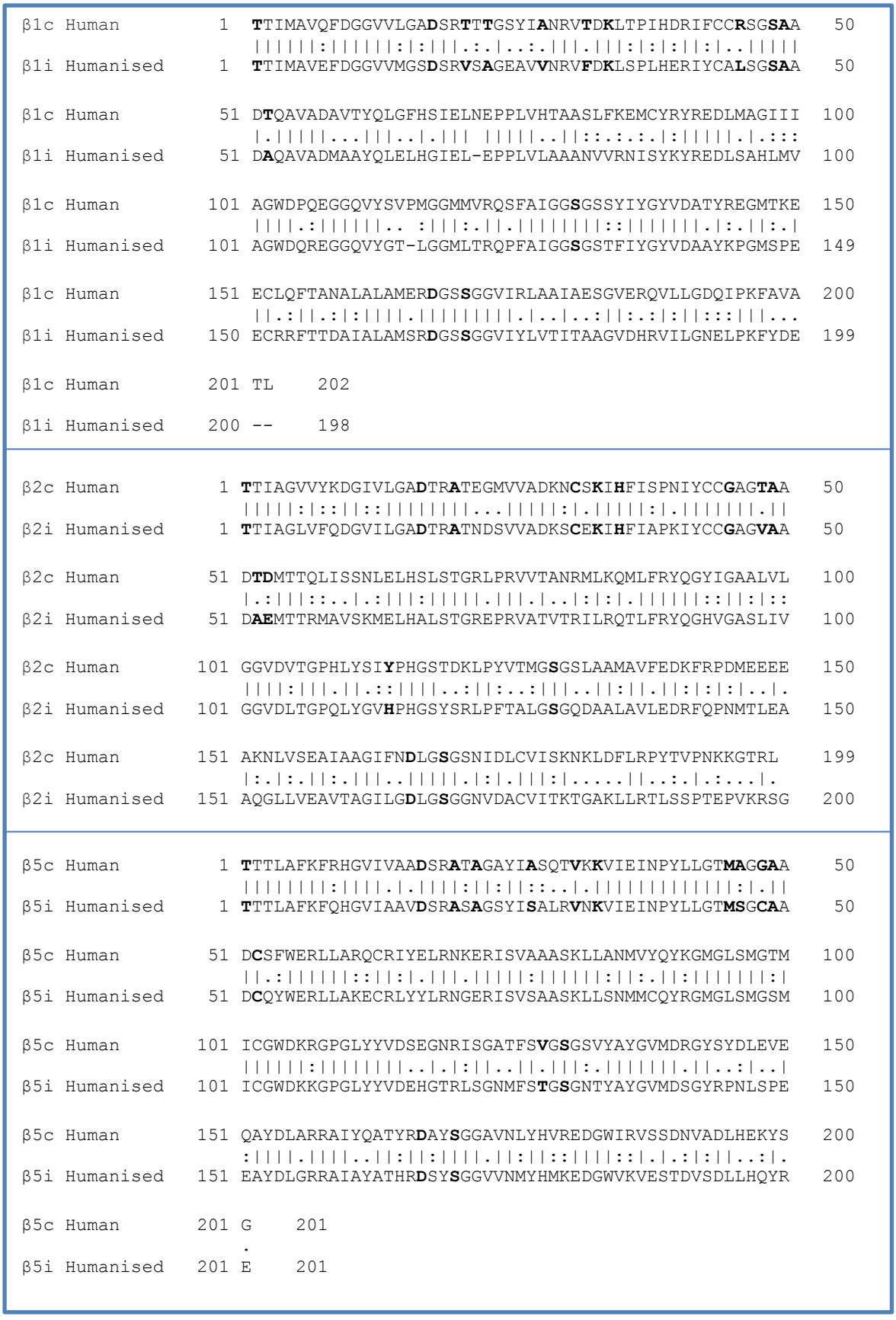
Score	Expect	Method	Identities	Positives	Gaps
380 bits(976)	4e-127	Compositional matrix adjust.	182/201(91%)	192/201(95%)	4/201(1%)

```
Query  59   TTTLAFKFQHGVIVAVDSRATAGSYISSLRMNKVIEINPYLLGTMSGCAADCQYWERLLA  118
            TTTLAFKFQHGVI AVDSRA+AGSYIS+LR+NKVIEINPYLLGTMSGCAADCQYWERLLA
Sbjct  69   TTTLAFKFQHGVIAAVDSRASAGSYISALRVNKVIEINPYLLGTMSGCAADCQYWERLLA  128

Query  119  KECRLYYLR----ISVSAASKLLSNMMLQYRGMGLSMGSMICGWDKKGPGLYYVDDNGTR  174
            KECRLYYLR    ISVSAASKLLSNMM QYRGMGLSMGSMICGWDKKGPGLYYVD++GTR
Sbjct  129  KECRLYYLRNGERISVSAASKLLSNMMCQYRGMGLSMGSMICGWDKKGPGLYYVDEHGTR  188

Query  175  LSGQMFSTGSGNTYAYGVMDSGYRQDLSPEEAYDLGRRAIAYATHRDNYSGGVVNMYHMK  234
            LSG MFSTGSGNTYAYGVMDSGYR +LSPEEAYDLGRRAIAYATHRD+YSGGVVNMYHMK
Sbjct  189  LSGNMFSTGSGNTYAYGVMDSGYRPNLSPEEAYDLGRRAIAYATHRDSYSGGVVNMYHMK  248

Query  235  EDGWVKVESSDVSDLLYKYGE  255
            EDGWVKVES+DVSDLL++Y E
Sbjct  249  EDGWVKVESTDVSDLLHQYRE  269
```

Supplementary Figure 11 - BLASTp alignment from murine β5i active site cut (query), to human sequence β5i (subject) – Murine IP sequence from PDB:3UNH, human β1i sequence from UniProt:P28062.

Supplementary table 9 - Substitutions performed in humanisation of IP active site structural data - Mutations performed using swissPDB viewer, from 20S Mouse IP PDB:3UNF active site cuts aligned to human LMP2 (P28065) for β1i, PSMB10 (P40306) for β2i, PSMB8 (P28062) for β5i sequences. '+' denotes conservative subsititutions.

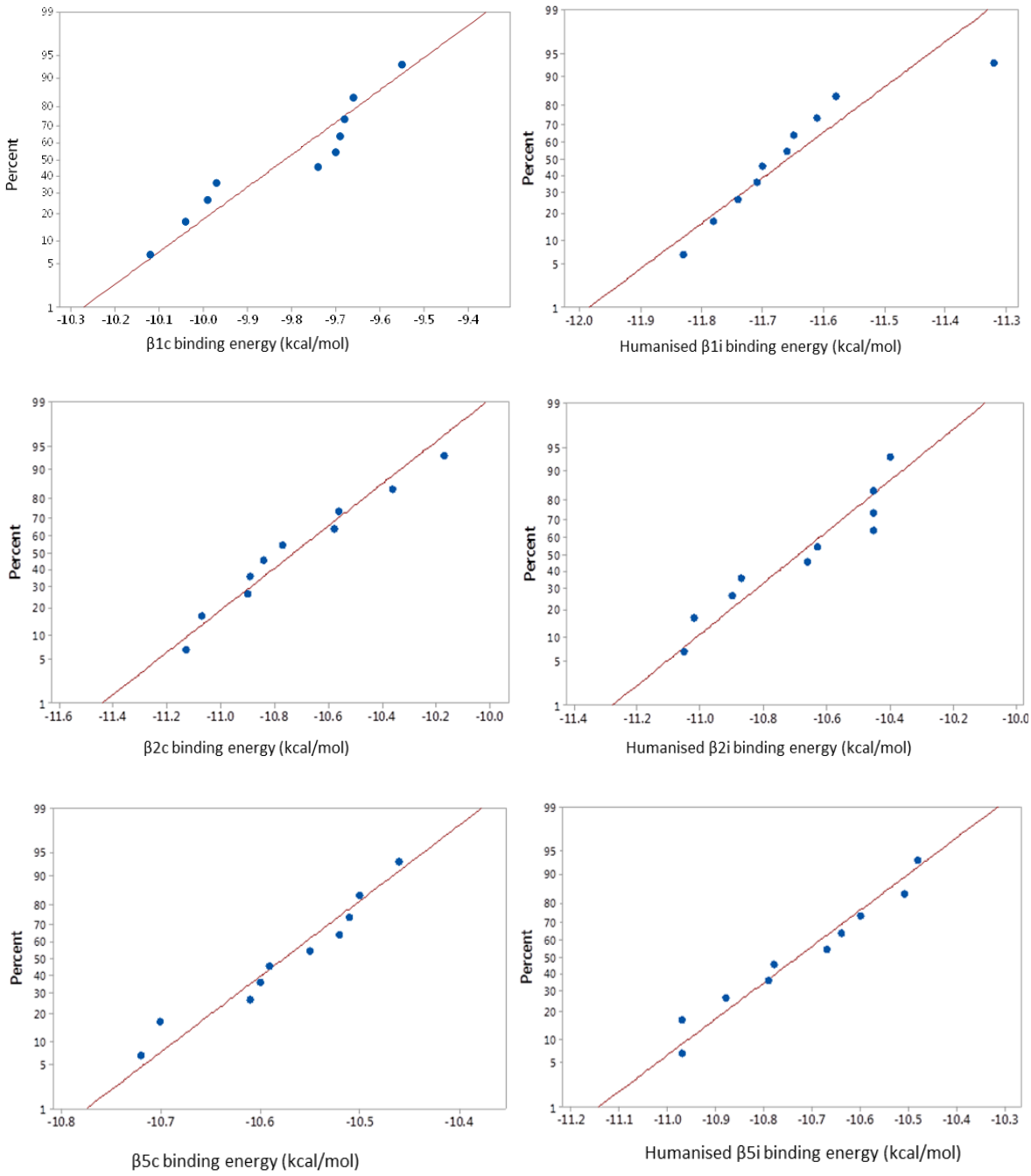
β1i Mouse		β1i Human	β2i Mouse		β2i Human	β5i Mouse		β5i Human
Val14	+	Met14	ARG 9	+	GLN9	VAL14		ALA14
Thr24		Glu24	THR52		ALA52	THR21	+	SER21
Gly39	+	Glu39	ALA60		VAL60	SER28	+	ALA28
Phe42	+	Tyr42	VAL99	+	ILE99	MET31	+	VAL31
Ile55	+	Val55	ASN106		THR106	LEU88		CYS88
Leu68	+	Ile68	GLU112		GLY112	ASP116	+	GLU116
Lys84	+	Arg84	GLY132		ASP132	ASN117	+	HIS117
Leu95		Ser95	VAL135	+	LEU135	GLN124		ASN124
Ile99	+	Met99	LEU137	+	VAL137	GLN145		PRO145
Met115	+	Leu115	GLU153		GLY153	ASP146	+	ASN146
Ile120		Thr120	ILE159	+	VAL159	ASN168	+	SER168
Thr125		Ala125	SER165		GLY165	SER190	+	THR190
Ser132	+	Thr132	ALA180		LYS180	TYR197	+	HIS197
Tyr133	+	Phe133	GLY181		THR181	LYS198	+	GLN198
Thr147	+	Ser147	GLN186		LEU186	GLY201		ARG201
Asn157	+	Asp157	ALA188		THR188			
Thr160		Ala160	THR191	+	SER191			
Asn164	+	Ser164	GLN197	+	LYS197			
Asp191	+	Asn191	ALA199	+	SER199			



Supplementary figure 12- Sequence alignment of human CP and humanised IP active site chain cuts – Performed using pBLAST, '|' = full conservation of the same residue, ':' = amino acids with strongly similar properties and '.' = weakly similar properties.

Supplementary table 10 - Raw data lowest binding energy from AutoDock simulations of argyrin B at CP and humanised IP sites – Lowest binding energy selected from 50 conformations over 10 replicates.

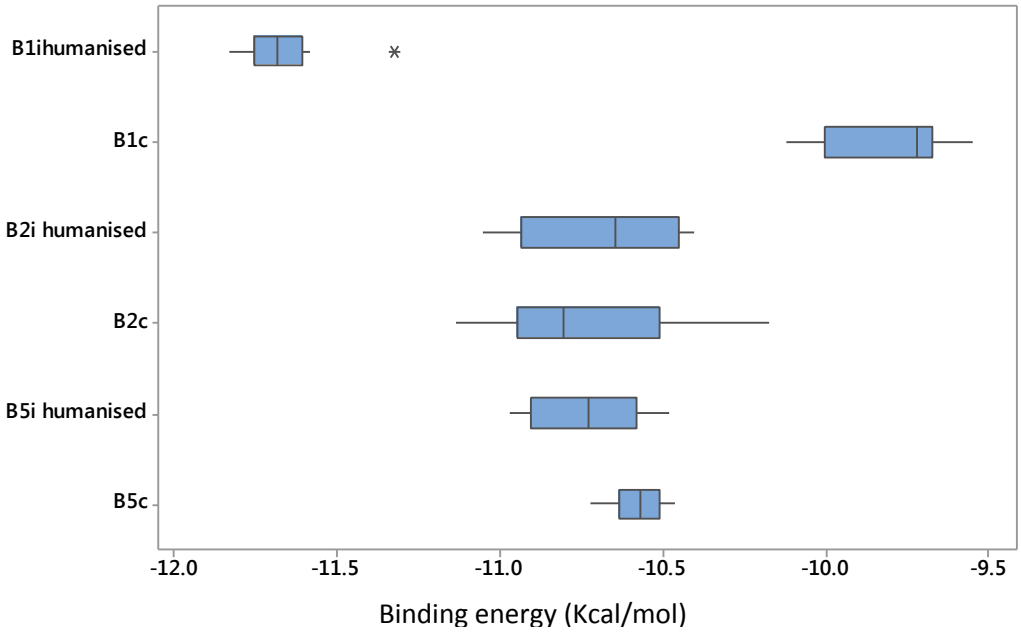
Lowest binding energy (Kcal/mol)					
β1ihumanised	β1c	β2i humanised	β2c	β5i humanised	β5c
-11.78	-9.99	-10.45	-10.84	-10.97	-10.52
-11.32	-9.55	-10.40	-10.56	-10.79	-10.51
-11.71	-9.70	-11.02	-10.17	-10.88	-10.61
-11.74	-9.68	-10.87	-10.58	-10.48	-10.72
-11.61	-9.74	-10.90	-11.07	-10.51	-10.59
-11.70	-9.69	-11.05	-10.90	-10.97	-10.60
-11.83	-9.97	-10.66	-10.77	-10.78	-10.55
-11.65	-9.66	-10.45	-10.89	-10.60	-10.50
-11.58	-10.12	-10.63	-10.36	-10.64	-10.46
-11.66	-10.04	-10.45	-11.13	-10.67	-10.70



Supplementary Figure 13 - Normality test probability plots from AutoDock simulation lowest binding energies of argyrin B at each CP and humanised IP active site – Using lowest binding energy from each of 10 repeats, analysis performed on Minitab 17.

Supplementary table 11 - Normality p-values on AutoDock lowest binding energy repeats at each CP and humanised IP active site - Calculated using Minitab 17 where if p>0.05, normal distribution is assumed.

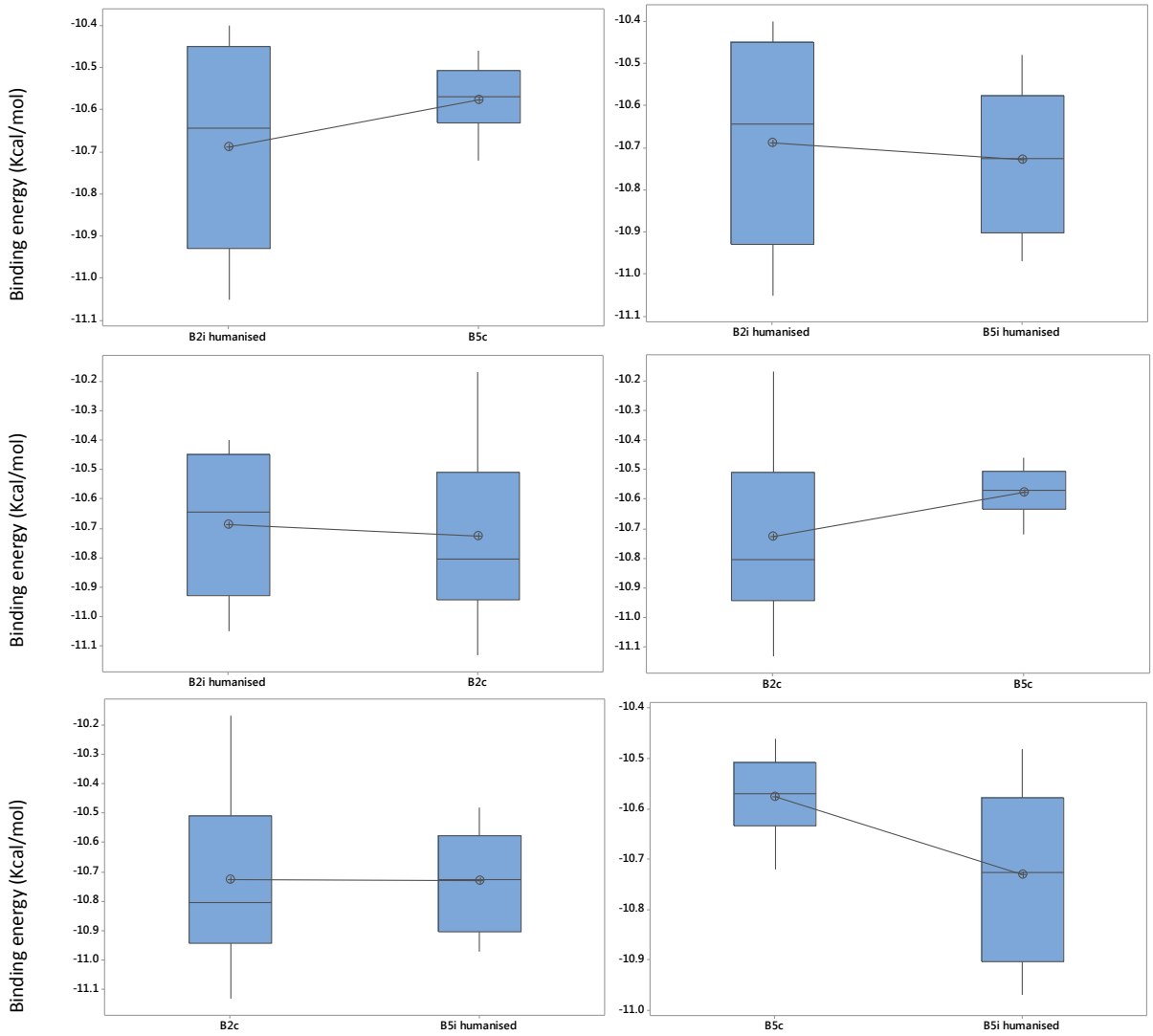
Binding energy (Kcal/mol)	β1c	β1i	β2c	β2i	β5c	β5i
Normality p-value	0.093	0.151	0.701	0.131	0.528	0.703
Mean	9.81	-11.66	-10.73	-10.69	-10.58	-10.73
SD	0.1959	0.1408	0.3065	0.2531	0.0853	0.1774



Supplementary Figure 14 - Test for equal variance on AutoDock lowest binding energy repeats of argyrin B at each CP and humanised IP active site.

Supplementary table 12 – Qualitative results for equal variance between AutoDock lowest binding energies of argyrin B at CP and humanised IP active sites - Y = equal variance, N = not of equal variance. Determines subsequent significance tests: Y = 2 sample t-test, N = non parametric Mann-Whitney test.

	β1c	β1i	β2c	β2i	β5c	β5i
β1c		N	N	N	N	N
β1i	N		N	N	N	N
β2c	N	N		Y	Y	Y
β2i	N	N	Y		Y	Y
β5c	N	N	Y	Y		Y
β5i	N	N	Y	Y	Y	



Supplementary Figure 15 - Boxplots displaying 2 sample t-test results between AutoDock lowest binding energies of argyin B at CP and humanised IP active sites - Data of normal distribution and equal variance.

Supplementary table 13 - 2 sample t-test p-values between AutoDock lowest binding energies of argyrin B at CP and humanised IP active sites – Data of normal distribution and equal variance. Analysis performed using Minitab 17 where, >0.05 = no statistical significance and <0.05 = significant difference.

	β1c	β1i	β2c	β2i	β5c	β5i
β1c		-	-	-	-	-
β1i	-		-	-	-	-
β2c	-	-		0.760	0.164	0.986
β2i	-	-	0.760		0.212	0.680
β5c	-	-	0.164	0.212		0.030
β5i	-	-	0.986	0.680	0.030	

Supplementary table 14 - Non-parametric Mann-Whitney tests p-values between AutoDock lowest binding energy of argyrin B at CP and humanised IP sites – Using data of normal distribution and non-equal variance. Analysis using Minitab where p>0.05 = no statistical significance and p<0.05 is of statistical significance.

	β1c	β1i	β2c	β2i	β5c	β5i
β1c		0.0002	0.0002	0.0002	0.0002	0.0002
β1i	0.0002		0.0002	0.0002	0.0002	0.0002
β2c	0.0002	0.0002		-	-	-
β2i	0.0002	0.0002	-		-	-
β5c	0.0002	0.0002	-	-		-
β5i	0.0002	0.0002	-	-	-	

Supplementary table 15 - AutoDock lowest binding energies for TMC-95A at yeast β2c – Upper quartile of best conformations selected from 5 repeats each of 50 conformations Bold = similar pose position to crystal structure data.

TMC-95A yeast β2c Lowest binding energy (Kcal/mol)				
1	**2**	**3**	**4**	**5**
-8.94	-8.70	-8.77	-8.81	-9.86
-8.55	-8.53	-8.74	-8.80	-8.52
-8.45	**-8.51**	-8.74	-8.73	-8.67
-8.38	-8.29	-8.38	**-8.57**	-8.66
-8.27	-8.24	-8.19	**-8.45**	-8.45
-8.14	**-8.15**	**-8.19**	**-8.34**	**-8.17**
-8.08	-8.12	**-8.18**	-8.30	**-8.03**
-8.03	-8.07	**-7.99**	-8.26	-7.99
-7.99	-8.02	-7.93	-8.15	-7.93
-7.97	**-8.00**	**-7.92**	**-8.14**	-7.92
-7.93	**-7.99**		-8.09	-7.92
	-7.99		-8.08	
	-7.96		-8.05	
	-7.96		**-7.97**	
	-7.95		**-7.95**	
			-7.95	

Equation 1.

$$v = \frac{d[P]}{dt} = k_2[ES]$$

Velocity is dependent on k_2 and concentration of ES complex.

Equation 2.

$$\frac{d[ES]}{dt} = k_1[E][S] - (k_{-1} + k_2)[ES]$$

The rate of change in the concentration of the enzyme-substrate complex is determined by the formation of ES subtract the breakdown of ES, either by the forward reaction to P + E, or reverse to E + S.

Equation 3.

$$0 = k_1[E][S] - (k_{-1} + k_2)[ES]$$

$$k_1[E][S] = (k_{-1} + k_2)[ES]$$

Initial velocities are measured and a steady state assumption is used. By using a low [E], this ensures that a significant proportion of the reaction was under the assumption that [ES] is constant, since [ES] will not significantly change unless a high [S] has been consumed. Therefore d[ES]/dt = 0, along with rearrangement forms equation 3.

Equation 4.

$$E_0 = [E] + [ES]$$

$$[E] = [E_0] - [ES]$$

At the start of the reaction the total [E] is the free E plus ES complex; see equation 4, including rearrangement.

Equation 5.

$$k_1([E_0] - [ES])[S] = (k_{-1} + k_2)[ES]$$

$$k_1[E_0][S] - k_1[ES][S] = (k_{-1} + k_2)[ES]$$

$$k_1[E_0][S] = (k_{-1} + k_2)[ES] + k_1[ES][S]$$

$$k_1[E_0][S] = (k_1[S] + k_{-1} + k_2)[ES]$$

$$[ES] = \frac{k_1[E_0][S]}{k_1[S] + k_{-1} + k_2}$$

Substituting equation 4 into 3 gives equation 5. Following this, the brackets are multiplied out, ES was arranged onto a single side and subsequently made the factor of the equation that is finally arranged alone.

Equation 6.

$$[ES] = \frac{[E_0][S]}{\frac{(k_{-1} + k_2)}{k_1} + [S]}$$

For simplification, dividing the right hand side numerator and denominator by k_1 produces equation 6.

Equation 7.

$$K_m = \frac{k_{-1} + k_2}{k_1}$$

From rearrangement of equation 3, [E][S]/[ES] has been termed as the Michaelis-Menten constant (Km), that produces equation 7.

Equation 8.

$$[ES] = \frac{[E_0][S]}{k_m + [S]}$$

With this in common, substituting equation 7 into 6 creates equation 8.

Equation 9.

$$k_2[ES] = \frac{k_2 [E_0][S]}{k_m + [S]} = v$$

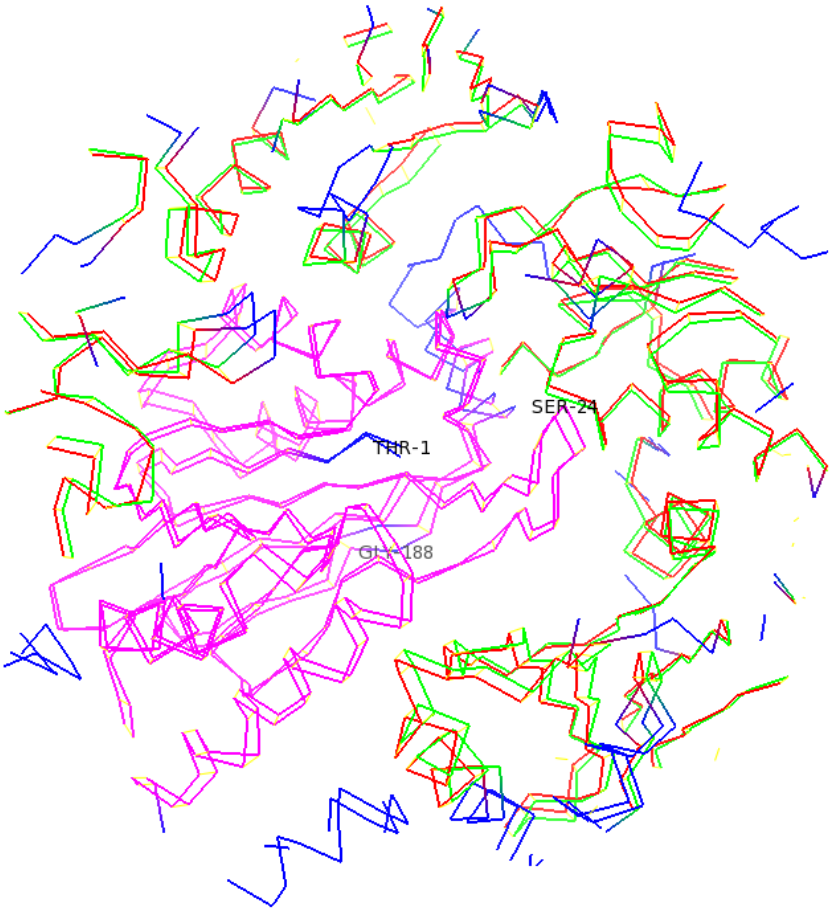
Combining equation 1 and 8, then multiplying both sides by k_2 for simplification, produces equation 9.

Equation 10.

$$v = \frac{V_{max}[S]}{K_m + [S]}$$

The highest initial rate, termed V_{max}, occurs when all enzyme is saturated in ES complexes. At this stage, [ES] = total [E]. Therefore, $k2[E_0] = Vmax$. Introducing this factor creates the final Michaelis-Menten equation 10.

Supplementary figure 16 - Derivation of Michelis-Menten equation – Series of equations showing the basis of K_m and V_{max} constants as well as parameters that non-linear regression analysis is based upon. E = enzyme, S = substrate, ES = enzyme-substrate complex, P = product, k = rate constant, d = 'difference in', v = velocity, K_m = Michaelis constant, V_{max} = maximal velocity.



Supplementary figure 17 - Humanised β5i aligned to chimeric yeast proteasome with human β5i subunit – Humanised β5i (red) created from mouse IP (PDB:3UNH) with substituted human β5i sequence (UniProt:P28062). Yeast proteasome contains human β5i (green) residues 1-138 and human B6 residues 97-11, 118-133, (PDB:5M2B). β5 chains shown in magenta and differences presented in blue. Interacting residues are cut within 28 Å of Thr1.

Ethical Approval

From: Stephanie Bee
Sent: 12 November 2015 11:04
To: Erika Loizidou
Cc: Adam Choonara; Duncan Allardyce
Subject: NSESC Approval for project no 1887 Duncan Allardyce "Determination of binding activity and intermolecular interactions between Argyrin B and the immunoproteasome by enzymatic assays and molecular modelling"

Dear Erika

Re: NSESC Approval for project no 1887 Duncan Allardyce "Determination of binding activity and intermolecular interactions between Argyrin B and the immunoproteasome by enzymatic assays and molecular modelling"

Please accept this email as confirmation that this project has been approved. Date of NSESC approval 10/11/2015.

Please note that the committee must be kept informed of any proposed changes to the project protocol.

Please advise the applicant to include a copy of this email in their final project submission as confirmation of ethical approval.

Adam Choonara

Chair of NSESC